

SEMIPARAMETRIC METHODS AND APPLICATIONS

A Dissertation

by

ZHONGJIAN LIN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Qi Li
Co-Chair of Committee,	Venkatesh Shankar
Committee Members,	Steven Puller
	Keli Xu
Head of Department,	Timothy Gronberg

May 2014

Major Subject: Economics

Copyright 2014 Zhongjian Lin

ABSTRACT

This dissertation studies semiparametric methods and their applications to economics and marketing problems. We propose a game theoretic model to analyze interactions among individuals in a social network with hierarchy. We find significant asymmetric peer effects among individuals in social networks. High status individuals deliver stronger peer effects on low status individuals than vice versa. Additionally, we investigate semiparametric panel data truncated regression models with fixed effects. We show the identification of our model with primitive assumption, establish the consistency and asymptotic normality of our proposed sieve estimator. We conclude that we can achieve \sqrt{n} -convergent rate for parametric parameters. Besides theoretical semiparametric methods, we study the dynamic effectiveness of marketing mix variables and the competition among the pioneer and early followers in pharmaceutical industry. With two pharmaceutical categories data, we find dynamic effectiveness of advertising and detailing inputs. Pioneer firms and follower firms have different effectiveness of advertising and detailing inputs in different stages. Our out-of-sample analyses show that when the data is rich, the semiparametric model outperforms the parametric model while it is better to deploy parametric model when the sample size is small.

ACKNOWLEDGEMENTS

I offer my sincerest gratitude towards my advisor Professor Qi Li for his continuous encourage, support and advice over the past five years. I am also indebted to all of my committee members: Professor Venkatesh Shankar, Professor Steven Puller and Professor Keli Xu for their constant guidance and encouragement. Their knowledge and personality have deeply shaped my traits as a young scholar. In addition, I am grateful to Professor Haiqing Xu for his tremendous help on my research.

Special recognition goes out to my family, for their support, encouragement and patience during my pursuit of the Doctorate in economics. My wife, Wenxin Guo, provided strong support and made many sacrifices. My parents continue to provide support of my pursuit of knowledge and degrees over the past twenty years. I am also thankful to my grandparents and my parents-in-law for their support of my study. This work is dedicated to them.

TABLE OF CONTENTS

	Page
ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iii
TABLE OF CONTENTS.....	iv
LIST OF FIGURES.....	vii
LIST OF TABLES.....	viii
CHAPTER I INTRODUCTION	1
CHAPTER II IDENTIFICATION OF HIERARCHY EFFECTS IN SOCIAL INTERACTIONS, WITH HAIQING XU	2
2.1 Introduction	2
2.2. Social Network Model	7
2.2.1. Status and Hierarchy Effects	8
2.2.2. Direct and Indirect Peer Effects	10
2.3. Characterize Bayesian Nash Equilibrium	11
2.4. Identification	14
2.5. Estimation.....	17
2.5.1. Maximum Likelihood Estimation	19
2.5.2. Nested Pseudo Likelihood Estimation	21
2.5.3. Consistency	23
2.5.4. Asymptotic Normality.....	24
2.5.5. Simulation-Based Estimation of Variance-Covariance Matrix.....	26
2.5.6. Testing Hierarchy Effects.....	27
2.6. Monte Carlo Experiments	27
2.7. College Attendance	31
2.7.1. Add Health Dataset	32
2.7.2. Peer Effects in College Attendance.....	34
2.8. Conclusion.....	36
CHAPTER III SEMIPARAMETRIC PANEL DATA TRUNCATED REGRESSION MODEL WITH FIXED EFFECTS, WITH CHUNRONG AI AND HONGJUN LI	37
3.1. Introduction	37
3.2. Panel Data Truncated Regression Model With Fixed Effects.....	41
3.3. Identification	43

3.4. Consistency	45
3.5. Asymptotic Normality	48
3.6. Conclusion.....	51
CHAPTER IV A SEMIPARAMETRIC APPROACH TO ESTIMATE THE DYNAMIC EFFECTIVENESS OF MARKETING MIX IN MARKET RESPONSE MODELS IN PIONEER-FOLLOWER CONTEXT, WITH VENKATESH SHANKAR.....	53
4.1. Introduction	53
4.2. Data	59
4.3. Determinants of Responsiveness.....	60
4.3.1. Order of Entry	60
4.3.2. Number of Competitors.....	61
4.3.3. Time in the Market.....	61
4.3.4. Network Size.....	61
4.3.5. Innovation.....	62
4.4. Model Development and Estimation.....	62
4.4.1. Parametric Multiplicative Model Specification	62
4.4.2. Smooth Varying Coefficient Model Specification.....	67
4.4.3. Model Estimation	68
4.5. Results	68
4.5.1. Estimation Results.....	68
4.5.2. Out-of-Sample Analysis	72
4.6. Managerial Implications.....	80
4.6.1. Implications Based on the Parametric Multiplicative Model Estimation ...	80
4.6.2. Implications Based on Smooth Varying Coefficient Model Estimation.....	80
4.7. Conclusion.....	81
CHAPTER V CONCLUSIONS.....	83
REFERENCES.....	84
APPENDIX	93

LIST OF FIGURES

	Page
Figure 1 Social Network and Social Status	6
Figure 2 Hierarchy Effects	10
Figure 3 Individual j Has Indirect Effect on Individual i Through Individual k.....	11
Figure 4 Smooth Varying Responsiveness of Fluoxetine Category	70
Figure 5 Smooth Varying Responsiveness of Statin Category	71
Figure 6 Model Fitting Comparison of Fluoxetine Category.....	73
Figure 7 Model Fitting Comparison of Statin Category	74
Figure 8 Out-of-Sample Predictions with Three to Five Years “In Sample” Time Windows: Fluoxetine	76
Figure 9 Out-of-Sample Predictions with Six to Seven Years “In Sample” Time Windows: Fluoxetine	77
Figure 10 Out-of-Sample Predictions with Three to Five Years “In Sample” Time Windows: Statin	78
Figure 11 Out-of-Sample Predictions with Six to Seven Years “In Sample” Time Windows: Statin	79
Figure 12 Social Distance and Neighborhood.....	96

LIST OF TABLES

	Page
Table 1 Mean and Standard Deviation I.....	30
Table 2 Mean Square Error I.....	30
Table 3 Mean and Standard Deviation II	30
Table 4 Mean Square Error II	31
Table 5 Mean and Standard Deviation III.....	31
Table 6 Mean Square Error III	31
Table 7 Summary of Statistics	33
Table 8 Estimation Results.....	36
Table 9 Literature on Market Response Models	57
Table 10 Comparison of Parametric and Semiparametric Models.....	58
Table 11 Summary Statistics of Key Variables	59
Table 12 Correlation Matrix for Model (1): Fluoxetine.....	65
Table 13 Correlation Matrix of Model (1): Statin.....	66
Table 14 Estimation Results: Fluoxetine	69
Table 15 Estimation Results: Statin	69
Table 16 Square of Prediction Errors with Different Time Windows	74

CHAPTER I

INTRODUCTION

In this dissertation, we study semiparametric methods and their empirical applications. The theoretical semiparametric methods range from discrete game model with sociological content of hierarchy and status; and panel data truncated model with fixed effects. Empirical applications include peer effects in college attendance decision among high school students with Add Health dataset; and dynamic effectiveness analysis of advertising and detailing in pharmaceutical industry. The theoretical studies enrich the literature in terms of combination of game model, econometrics and sociology and extend traditional parametric panel data truncated model with fixed effects to allow for flexible nonparametric components. Empirical study in college attendance provides policy implication that government can target on high status student to increase college enrollment rate. Empirical study in oligopolistic competition points out that the effectiveness of advertising and detailing are different for pioneer and follower firms and therefore the managerial implications from this study are different for pioneer and followers firms in different stages of competition.

CHAPTER II

IDENTIFICATION OF HIERARCHY EFFECTS IN SOCIAL INTERACTIONS,
WITH HAIQING XU

2.1 INTRODUCTION

Network models in the recent literature using game theoretic frameworks have been much successful to understand social interactions that are traditionally discussed in sociology. A leading example is the network formation, e.g. Jackson and Wolinsky (1996); Bala and Goyal (2000) for the theory side and Christakis et al. (2010); Sheng (2012); Mele (2011); Badev (2013) through the structural and empirical ways. Another example considers the network interactions, e.g., Jackson and Yariv (2007) and Galeotti et al. (2010) for theoretic analysis and Brock and Durlauf (2001); Xu (2011); Kline (2012) from econometric methodological perspective. The empirical applications range over the peer effects in education (e.g. Sacerdote, 2001; Angrist and Lang, 2004; Calvó-Armengol et al., 2009), labor economics (e.g. Duflo and Saez, 2003), health issue (e.g. Trogdon et al., 2008; Carrell et al., 2011), juvenile behavior (e.g. Powell et al., 2005; Bayer et al., 2009; Patacchini and Zenou, 2012) and technological progress of a country (e.g. Fogli and Veldkamp, 2013) to name only a few.

In this paper, we focus on the hierarchy effects in a social network. In sociology, hierarchy is a system in which people are divided into different levels of social status. In a social network, interactions between a pair of individuals are quite sensitive to their relative social status. Ball et al. (2001) use experiment approach and find that that higher status individuals in general are more influential and obtain more surplus than the same

individuals but with lower status. Bault et al. (2008) use experiment approach to study how Social Ranking Affects Choice Behavior. Ballester et al. (2006) establish a Bonacich (centrality)-Nash (equilibrium) linkage by looking at the role of key players in the social network and Becker et al. (2005) study the role of social status played in their resource allocation problem and the inequality. Another line of research, e.g. Robson (1992); Akerlof (1997); Luttmer (2005); Ray and Robson (2012) focus on how the relative social status directly affect the utilities of agents. For more details on the hierarchy and status, we refer to Wasserman and Faust (1994); Podolny (2005); Sauder, Lynn and Podolny (2012), etc.

Our paper contributes to the empirical social interaction literature that emphasizes the role of network in explaining direct and indirect peer effects and the strength of the indirect effects (Xu, 2011; Blume et al., 2013). First, we develop a game model of incomplete information to capture the interactions among individuals with different social status in the network. The utility function of each individual depends not only on own choice but also her friends' choices. The strength of such interactions relies on the social status of her friends as well as her own status. For example, the behavior of an individual with a low social status could be influenced by a high status friend, but the effect would be much smaller if this friend also comes from the low social status group. For policy purposes, it is our interest to define and quantify the hierarchy effects, which is clearly different from the peer effects. In our framework, both effects can be identified and distinguished from each other. Our model is related to Xu (2011), which investigates only peer effects.

A key advantage of our game theoretic approach is to deal with the simultaneity of interactions in social interactions by following the standard Bayesian Nash Equilibrium (BNE) solution concept. The simultaneity issue has long been central to sociology and economics (see e.g. Manski, 1993, the “reflection problem”). Ignoring such an issue would cause endogeneity in empirical analysis. In equilibrium, strategic effects are pervasive in a social network: an individual’s behavior is affected by her friends, her friends’ choices are further affected by friends of friends, etc. Along the network, one individual affects the others directly or indirectly. Such simultaneity is well captured by the BNE solution concept. Similar as the work by Brock and Durlauf (2001); Durlauf and Ioannides (2010); Blume et al. (2011) and Xu (2011), we employ a static discrete game of incomplete information.

Second, we establish the identification and estimation of the proposed game model. Under primitive conditions, we provide semiparametric identification of the structural game model. In particular, we require the number of friends to be bounded above and the interaction strength to be reasonably small. The former is needed for the reason that the data actually come from the equilibrium of one single large game, instead of repetitions of the same game. A similar restriction is implicitly embedded in time series where the data are from a single time path. The latter condition is analogous to the unit root restriction in time series for the dependent data analysis. This condition is imposed for two reasons. First, it ensures that the equilibrium is unique. Under a similar condition, Brock and Durlauf (2001) also obtain the uniqueness of equilibrium in their setting. In empirical analysis, an obvious obstacle from multiple equilibria is the incompleteness of

the econometric model (see Tamer, 2003). The second reason is more essential: the small interaction strength implies the network stability condition (see e.g. Jackson and Wolinsky, 1996; Xu, 2011), which limits the dependence of choices across individuals in the network.

For estimation, we propose a nested pseudo likelihood estimation (NPLE) method for our large game with social interactions. In a seminal paper by Aguirregabiria and Mira (2007), they propose NPLE for solving the computational difficulty in dynamic games. In particular, this approach does not require solving the fixed point introduced by the definition of the equilibrium. It is a natural extension of their approach to our large game: similarly to dynamic games, it is costly to compute the equilibrium in a large network game using a fixed point algorithm. This is because the space that the fixed point lives in has a high dimension, which equals to the number of individuals in the social network. The advantage of this algorithm is that we do not actually compute any equilibrium of the game, but adopt an iterative estimation procedure that converges. Note that our NPLE is related to the MLE type estimator suggested by Xu (2011), but the difference is essential.

Specifically, the proposed NPL method starts with an arbitrary guess of the choice probabilities, e.g. the predicted choice probabilities from a regular Logit estimation that takes no strategic effects into account. Then we conduct another Logit estimation by using the given choice probabilities as the equilibrium choice probabilities. After that, we obtain an update of the predicted choice probabilities from the estimates. So on and so forth, our iterative algorithm for NPLE consists of a sequence of Logit estimations

until the estimates converge and for each iteration, we update the equilibrium choice probabilities. In a large social network game, the NPLE is quite attractive to practitioners due to its simplicity for implementation. Under further regularity conditions, we establish the consistency and asymptotic normality for our NPL estimator.

The rest of the paper is organized as follows. Section 2 gives an introduction of the social network model and the concepts of status and hierarchy. Section 3 characterizes the BNE and establishes its uniqueness. Section 4 provides nonparametric identification of our model. Section 5 introduces two estimation methods, Maximum Likelihood Estimation and Nested Pseudo Likelihood Estimation, and establishes their consistencies and asymptotic normality. Three Monte Carlo experiments are executed in Section 6. An empirical study of college attendance is conducted in section 7. All proofs are given in the appendix.

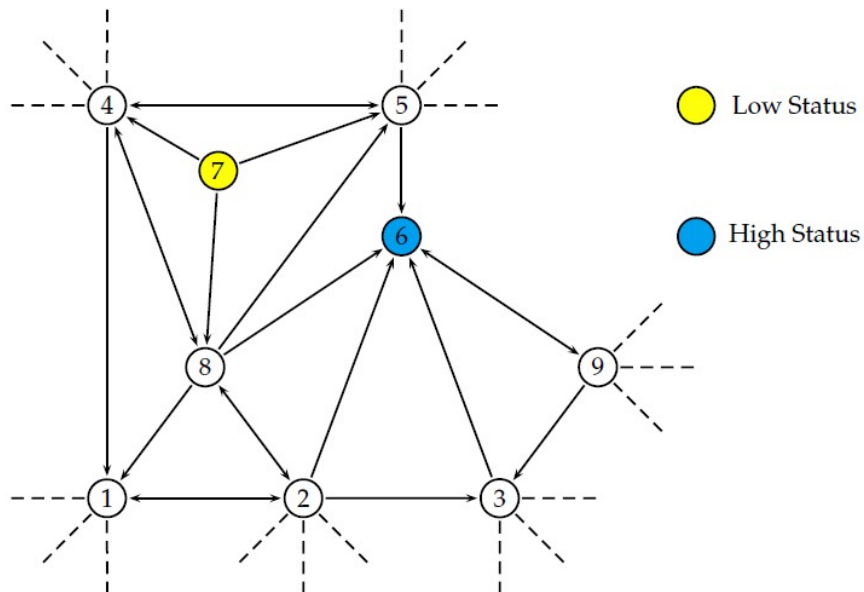


Figure 1 Social Network and Social Status

2.2. SOCIAL NETWORK MODEL

We consider a static discrete game of incomplete information in a large social network. There are n individuals, indexed by $i \in I \equiv \{1, \dots, n\}$, exogenously located in the network and each individual i is socially connected to a group of friends: i names a set of individuals as her friends, denoted as $F_i \subseteq I/\{i\}$. Let $\ell_{ij} = 1$ if $j \in F_i$ and $\ell_{ij} = 0$ otherwise. Therefore $F_i = \{j \neq i: \ell_{ij} = 1\}$. Note that the friendships need not be symmetric, i.e. $\ell_{ij} \neq \ell_{ji}$ is allowed. An example of social network is shown in Figure 1.

Each individual i simultaneously makes a binary choice $Y_i \in \{0, 1\}$. In our empirical study in section 7, we consider college attendance decisions of high school students. In particular, we take $Y_i = 1$ as the choice of enrolling for undergraduate education and $Y_i = 0$ for not entering colleges. The utility function of individual i choosing $Y_i = 1$ is given as follows:

$$U_{i1} = \beta(X_i) + \sum_{j \in F_i} \alpha(Y_j, S_i, S_j) + \epsilon_i \quad (1)$$

The first component is the mean utility that is determined by each individual's demographic characteristics $X_i \in \mathbb{R}^d$. The second term characterizes peer effects from the choices of individual i 's friends. Last, there is an unobserved utility shock ϵ_i . When individual i chooses $Y_i = 0$, the utility is set to be zero, i.e. $U_{i0} = 0$, which is a standard payoff normalization in binary response models.

In equation (1), S_i is the social status of individual i , which is the main objective of our interest. The social status is related to the respect one has in the eyes of others, Magee and Galinsky (2008). In our setting, the social status is determined by the number

of friend nominations received from others. To illustrate, consider the social network established by Twitter, we can take S_i as the number of followers. For notational simplicity, we take S_i as a binary variable in our empirical analysis. Specifically, $S_i = H$ if individual i is with a “high” social status and $S_i = L$ if she is with a “low” social status. The high or low social status is determined by the number of nominations from other individuals in the network and captures the centrality of individuals in the network. The functions α and β are structural parameters of interest, which are nonparametric. Not using parametric assumptions on the utility function is practically cumbersome, but ensures that the identification of our model comes from the essential model restrictions, instead of from a particular selection of the parametric function form. In our empirical study, however, we assume $\beta(\cdot)$ is a linear index which can be estimated at the \sqrt{n} rate. It is worth pointing out that the network structure is assumed to be exogenously given in our model and we consider static game. The evolution of network formation is an interesting and deep subject in the literature, see, e.g., Christakis et al. (2010); Mele (2011); Sheng (2012); Kline (2012) and Badev (2013), however, which is not the focus of this paper.

2.2.1. Status and Hierarchy Effects

A crucial concept in our analysis is the social status, which mainly determines the strength of peer effects in a social network. Social status is a special variable that is quite different from other individual characteristics. Specifically, it captures important features of the social network, i.e., the centrality of an individual in the network. In this paper, the measure of social status depends on the number of friend nominations from others. A

more sophisticated measure of social status is the Katz-Bonacich centrality which accounts for the importance of an individual in the network, see Kats (1953); Bonacich (1987) for details. In the literature, social status plays an important role in the utility function. For example, Akerlof (1997) develops a status model in which status directly affects individuals' utilities and analyzes how hierarchy affects decision makings. Becker et al. (2005) take status as a "status good" in the marginal utility of consumption -the higher the status, the larger the marginal consumption utility. Ball et al. (2001) use an experimental design to study how social status affects bargaining payoffs and find that players with high status always obtain more surpluses. In our setting, social status affects the utility function through the interaction strength α . The interaction strength not only depends on her own social status, but also the status of her friends. In other words, the social interactions depend on the structure of the network and its hierarchy. Namely, hierarchy is the distribution of the social status in the network. Theoretically, peer effects should be monotonely increasing in friends' social status S_j but decreasing in own status S_i . In particular, social interactions are asymmetric, i.e., $\alpha(\cdot, S_i, S_j) \neq \alpha(\cdot, S_j, S_i)$ when $S_i \neq S_j$. Figure 2 illustrates the idea of hierarchy effects.

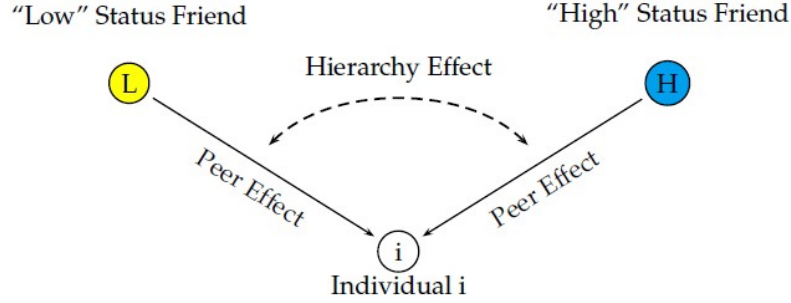


Figure 2 Hierarchy Effects

2.2.2. Direct and Indirect Peer Effects

In the specification of our utility function, we deploy the local interaction setting and the direct social interactions only occur among friends. However, the interactions could transmit through individual and the individuals can affect each other indirectly. For example, let $j \notin F_i$, but $k \in F_i$ and $j \in F_k$. Then the choice of individual j affects k 's utility, hence i needs to consider the possibility of j 's choice if i wants to take k 's choice into account. Given the structure of the social network, strategic effects pass through friends, friends of friends, etc., and finally, reach the whole network in equilibrium. An important property of such indirect strategic effects is that under weak and primitive conditions introduced later, these effects decay with the network distance, i.e. a condition called network stability condition (NSC), see Xu (2011). Under NSC, the dependence between any two individuals' choices vanishes as the network distance between them increases.

The indirect peer effects have been explicitly studied in the theoretic network literature, e.g. Jackson and Wolinsky (1996). In equilibrium, these indirect peer effects pervade the large network in local interaction models, e.g. Ellison (1993) and Özgür (2011). In our setting, the parameter of interest is: for $y = 0,1$ and $s, s' = L, H$,

$$\alpha(y, s, H) - \alpha(y, s, L) \text{ and } \alpha(1, s, s') - \alpha(0, s, s')$$

which quantify the hierarchy effects and peer effects, respectively. Figure 3 illustrates the ideas of indirect and direct effects.

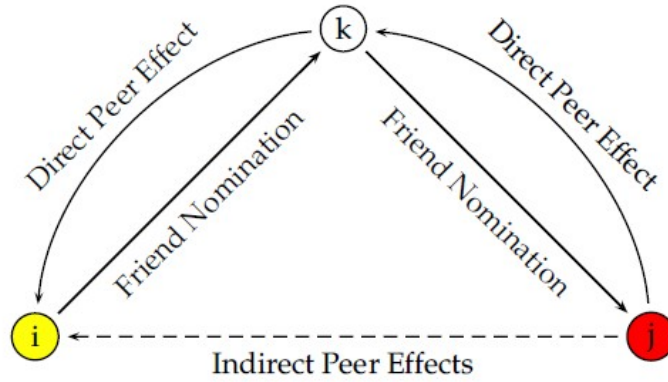


Figure 3 Individual j Has Indirect Effect on Individual i Through Individual k.

2.3. CHARACTERIZE BAYESIAN NASH EQUILIBRIUM

We now characterize the Bayesian Nash Equilibrium of the social network game where individuals make binary choices simultaneously. In particular we will show that BNE exists and is unique under weak conditions. More importantly, we will also establish a property of equilibrium strategies, denoted as “network stability”. Such a

property is important for our asymptotic analysis. To proceed, we first characterize the BNE solution to our game. We make the following assumptions for our discussion.

Assumption 1. *Let ϵ_i be i.i.d. across players and conform to the logistic distribution, i.e. $F_{\epsilon_i}(t) = e^t / (1 + e^t)$.*

Assumption 1 is standard in binary responses models. Similar assumptions can also be found in the empirical game literature, e.g. Brock and Durlauf (2001); Bajari et al. (2010) and Xu (2011). As we will see later, assumption 1 helps provide a closed form for choice probabilities in terms of best responses. However, this distributional assumption is not essential and our results can also generalize to other distributions in the exponential family, e.g., normal distribution.

Assumption 2. *Let $\max_{i \in I} \sum_j \ell_{ij} \leq M$, where M is a natural number and invariant with n .*

Assumption 2 requires that the number of friends to be less than a fixed number no matter how large is the social network, in particular, when the number of individuals in the network goes to infinity in our asymptotic analysis. By such a condition, we rule out the case that the network is highly centralized, e.g. “Star Network”. In our dataset from the National Longitudinal Study of Adolescent Health (Add Health), each student has at most 10 friends. Therefore $M = 10$ in our empirical study.

Assumption 3. *Let $\lambda = \max_{s, s'} |\alpha(1, s, s') - \alpha(0, s, s')| \times \frac{M}{4} < 1$.*

Assumption 3 amounts to the unit root restriction in time series and limit the strength of social interactions. With such a restriction, the dependence among

equilibrium decisions would satisfy the mixing conditions for dependent data. Similar assumptions can also be found in Brock and Durlauf (2001) and Xu (2011).

Let $W_n = \{(X_1, S_1, F_1), \dots, (X_n, S_n, F_n)\}$ be all the public information in the game and $\theta = (\alpha, \beta)$ be the parameters of interest. By the BNE solution concept, individual i 's equilibrium decision can be written as $Y_i = r_i^*(W_n, \epsilon_i)$, where r_i^* is the equilibrium strategy satisfying the mutual consistency conditions from the BNE solution concept. Namely,

$$r_i^*(W_n, \epsilon_i) = 1 \left\{ \beta(X_i) + \sum_{j \in F_i} \mathbb{E}[\alpha(r_j^*, S_i, S_j) | W_n, \epsilon_i] - \epsilon_i \geq 0 \setminus Big \right\}$$

Because r_i^* is binary valued, we have

$$\begin{aligned} \mathbb{E}[\alpha(r_j^*, S_i, S_j) | W_n, \epsilon_i] &= \mathbb{E}[\alpha(r_j^*, S_i, S_j) | W_n] \\ &= \alpha(1, S_i, S_j) \cdot P(r_j^*(W_n, \epsilon_j) = 1 | W_n) + \alpha(0, S_i, S_j) \\ &\quad \cdot P(r_j^*(W_n, \epsilon_j) = 0 | W_n) \end{aligned}$$

where the first step comes from the i.i.d. condition in assumption 1.

Denote $p_i(W_n; \theta) = P(r_j^*(W_n, \epsilon_j) = 1 | W_n)$. By assumption 1, we have

$$p_i(W_n; \theta) = \frac{\exp[\beta(X_i) + \sum_{j \in F_i} \alpha(0, S_i, S_j) + \sum_{j \in F_i} \gamma(S_i, S_j) \cdot p_j(W_n; \theta)]}{1 + \exp[\beta(X_i) + \sum_{j \in F_i} \alpha(0, S_i, S_j) + \sum_{j \in F_i} \gamma(S_i, S_j) \cdot p_j(W_n; \theta)]} \quad (2)$$

where $\gamma(S_i, S_j) = \alpha(1, S_i, S_j) - \alpha(0, S_i, S_j)$.

Let $P_n(W_n; \theta) = (p_1(W_n; \theta), \dots, p_n(W_n; \theta))'$. Further we define the best response function as follows: for each $p = (p_1, \dots, p_n) \in [0, 1]^n$,

$$\Gamma_i(p, W_n; \theta) = \frac{\exp[\beta(X_i) + \sum_{j \in F_i} \alpha(0, S_i, S_j) + \sum_{j \in F_i} \gamma(S_i, S_j) p_j]}{1 + \exp[\beta(X_i) + \sum_{j \in F_i} \alpha(0, S_i, S_j) + \sum_{j \in F_i} \gamma(S_i, S_j) p_j]}$$

Let further $\Gamma = (\Gamma_1, \dots, \Gamma_n)'$. Then a solution $P_n^*(W_n; \theta)$ to equation (2) can be rewritten as a fixed point of the following equation in $p \in [0,1]^n$:

$$p = \Gamma(p, W_n; \theta)$$

The existence of a solution to equation (2) has been guaranteed by the Brouwer's fixed point theorem; see, e.g., Bajari et al. (2010) for a similar result. Under assumptions 1 to 3, we can further obtain the uniqueness of the BNE solution.

Theorem 1. *Under assumptions 1 to 3, there exists a unique pure strategy BNE for any n .*

Proof. See Appendix B.1.

In this paper, we restrict ourselves to a unique BNE. Without assumptions 1 to 3, there may be multiple equilibria. In such a case, an obvious obstacle is the incompleteness of the econometrics model, i.e. each value of the structural parameter θ could deliver more than one observations of the decisions on the social network. Then what we observe in the data is some mixture of several equilibrium distributions of observables. For our network game with large number of players, it is difficulty to incorporate multiple equilibria and apply partial identification approach. For more discussions on this issue, see e.g. Tamer (2003).

2.4. IDENTIFICATION

In this section we consider semiparametric identification of the structural parameter θ . The definition of identification of the game model follows Hurwicz (1950) and Koopmans and Reiersøl (1950). Specifically, given the joint distribution of $F_{Y_1, \dots, Y_n | W_n}$,

can we uniquely derive the structural parameter θ ? If the answer is “yes”, then we obtain identification of the structural model; otherwise, the model is not identified.

Our identification is constructive, similar as that of Robinson (1988). Let $\Delta(W_n) = \ln \mathbb{E}[Y_i|W_n] - \ln\{1 - \mathbb{E}[Y_i|W_n]\}$ be the difference between the logarithm probabilities of choosing 1 and 0. Because $\Delta(W_n)$ is derived from $\mathbb{E}[Y_i|W_n]$ and then from the distribution $F_{Y_1, \dots, Y_n|W_n}$, we can take $\Delta(W_n)$ as a known object in our identification analysis.

Note that $p_i(W_n; \theta)$ can be identified by $\mathbb{E}[Y_i|W_n]$ through Theorem 1. Moreover, by equation (2), we have

$$\Delta(W_n) = \beta(X_i) + \sum_{j \in F_i} \{\alpha(0, S_i, S_j) + \gamma(S_i, S_j) \cdot p_j(W_n; \theta)\} \quad (3)$$

For notational simplicity, we discuss identification by focusing on the case where S_i is binary. In particular, S_i takes binary values: $S_i \in \{L, H\}$ where L and H represent the low and high social status position, respectively. Then, we can rewrite $\Delta(W_n)$ as

$$\Delta(W_n) = \beta(X_i) + \sum_{s \in \{L, H\}} \alpha(0, S_i, s) Z_{\alpha is} + \sum_{s \in \{L, H\}} \gamma(S_i, s) Z_{\gamma is} \quad (4)$$

where $Z_{\alpha is} = \sum_{j \in F_i} 1\{S_j = s\}$ and $Z_{\gamma is} = \sum_{j \in F_i} 1\{S_j = s\} \mathbb{E}[Y_j|W_n]$ for $s = L, H$. Note that $Z_{\gamma is}$ can also be obtained from the distribution of the observables, $F_{Y_1, \dots, Y_n|W_n}$ and $Z_{\alpha is}$ is simply a transformation of the observables. Fix $x \in \mathcal{S}_X$ and $s \in \{L, H\}$. For all $w \in \mathcal{S}_{W_n|X_i=x, S_i=s}$, note that equation (4) becomes

$$\mathbb{E}[\Delta(W_n)|W_n = w] = \beta(x) + \sum_{s' \in \{L, H\}} \{\alpha(0, s, s') z_{\alpha is} + \gamma(s, s') z_{\gamma is}\}$$

where $z_{\alpha i s'} = \sum_{j \in F_i} 1\{S_j = s'\}$ and $z_{\gamma i s'} = \sum_{j \in F_i} 1\{S_j = s'\} \mathbb{E}[Y_j | W_n]$. Therefore, when the support $\mathcal{S}_{W_n | X_i=x, S_i=s}$ is rich enough, we obtain the identification for the coefficients, $\beta(x)$, $\alpha(0, s, s')$ and $\gamma(s, s')$ as if in a linear model.

Formally, we summarize the above discussion in the following theorem. To begin with, let $Z_i = (1, Z_{\alpha i L}, Z_{\alpha i H}, Z_{\gamma i L}, Z_{\gamma i H})'$. We now impose the following rank condition on the support $\mathcal{S}_{W_n | X_i=x, S_i=s}$.

Assumption 4. (No multicollinearity) Let $\mathbb{E}[Z_i Z_i' | X_i = x, S_i = s]$ have the full rank for all $x \in \mathcal{S}_X$ and $s \in \{L, H\}$.

To satisfy assumption 4, (X_j, S_j) need to have variations while conditioning on X_i and S_i which effectively changes the equilibrium choice probabilities of individual i 's friends. Moreover, assumption 4 also implicitly requires that the number of friends (NF), i.e., $NF_i = \sum_{j \neq i} \ell_{ij}$, has variation. This is due to the fact that $Z_{\alpha i L} + Z_{\alpha i H} = NF_i$.

For $s \in \{L, H\}$, let $\alpha(s) = [\alpha(0, s, L), \alpha(0, s, H)]'$ and $\gamma(s) = [\gamma(s, L), \gamma(s, H)]'$. In the next theorem, we provide the identification of $(\alpha(s), \beta(x), \gamma(s))$. Because x and s are arbitrary, thus we establish the identification of θ .

Theorem 2. Suppose assumptions 1 to 4 hold. For any x and s , $\beta(x), \alpha(s)', \gamma(s)')$ is identified by

$$\beta(x), \alpha(s)', \gamma(s)')' = \{\mathbb{E}[Z_i Z_i' | X_i = x, S_i = s]\}^{-1} \mathbb{E}[Z_i \Delta(W_n) | X_i = x, S_i = s] \quad (5)$$

Proof. The proof directly follows our discussions above, and hence is omitted.

2.5. ESTIMATION

In this section, we propose a simple estimation procedure called nested pseudo likelihood approach, which is motivated from Aguirregabiria and Mira (2007). Intuitively, the difficulties in our setting come from the computational burden of solving the equilibrium of the large network game, which is quite similar to the dynamic game in Aguirregabiria and Mira (2007). Later, we illustrate this approach by using Monte Carlo experiments.

Consider a sample $\{(Y_i, X_i, F_i, S_i): i = 1, \dots, n\}$ from a large social network. Note that all the data come from only one game, instead of from repetitions of the same small game. Moreover, our asymptotic analysis relies on the number of individuals in the network going to infinity, which has a similar flavor to the analysis in time series and spatial econometrics. In empirical studies, clearly our approach can be applied to the case where all observations come from a small number of networks and each network contains a large number of individuals.

For our estimation, we employ a parametric linear-index setting:

$$U_{i1} = X_i' \beta + \sum_{j \in F_i} \{\alpha(0, S_i, S_j) + \gamma(S_i, S_j) p_j(W_n; \theta)\} - \epsilon_i \quad (6)$$

where $\beta \in \mathbb{R}^d$. Note that we can take this setting as a natural extension of the single-agent Logit model, i.e. $\alpha(y, s, s') = 0$ for all y, s and s' . Moreover, let

$$\begin{aligned} \gamma &= (\gamma(L, L), \gamma(L, H), \gamma(H, L), \gamma(H, H))', \\ \alpha &= (\alpha(0, L, L), \alpha(0, L, H), \alpha(0, H, L), \alpha(0, H, H))'. \end{aligned}$$

For notational simplicity, we denote $\theta_0 = (\alpha'_0, \beta'_0, \gamma'_0)' \in \Theta \subseteq \mathbb{R}^{d+d_s}$ as the underlying true parameter, and $\theta = (\alpha', \beta', \gamma')'$ as a generic parameter in Θ . In our setting of two status, $d_s = 8$. Recall that the main objective of our interest are the hierarchy effects and peer effects, which now become $\alpha_0(\cdot, \cdot, H) - \alpha_0(\cdot, \cdot, L)$ and γ_0 , respectively, in this parametric setup.

The unique equilibrium in Theorem 1 is crucial for our estimation analysis. In addition, we need to strengthen assumption 3 to rule out multiple equilibria for all Θ .

Assumption 5. For all $\theta \in \Theta$, let

$$\lambda = \max_{s, s' \in \{L, H\}} |\gamma(s, s')| \times \frac{M}{4} < 1$$

In the Add Health dataset, we have $M = 10$. Therefore, we need to restrict our parameter space for any analysis based on Add Health dataset by $\gamma(s, s') \in (-0.4, 0.4)$ for all s and s' .

Lemma 1. *Under assumptions 1, 2 and 5, the network game has a unique BNE for all n and $\theta \in \Theta$.*

The proof of Lemma 1 directly follows Theorem 1, and hence is omitted. For the purpose of comparison, here we first introduce the maximum likelihood estimator (MLE), which is the most efficient under regularity conditions, but impractical due to its heavy computational burden in large network games. In contrast, our NPLE is relatively easy to execute and its standard error can obtain in a direct manner.

2.5.1. Maximum Likelihood Estimation

Under the uniqueness of the equilibrium in Theorem 1, $P_n^*(W_n; \theta)$ is well defined. Now we are ready to define the likelihood function. For each $\theta \in \Theta$, let $L_i(Y_i, W_n; \theta) = Y_i \ln p_i^*(W_n; \theta) + (1 - Y_i) \ln 1 - p_i^*(W_n; \theta)$ be the log-likelihood of individual i 's choice (conditional on W_n). Note that the log-likelihood of observation i depends not only on (Y_i, X_i, S_i, F_i) , but also on (X_{-i}, S_{-i}, F_{-i}) , which is due to the strategic effects among individuals. Let

$$L(\theta; n) = \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n L_i(Y_i, W_n; \theta) \right\} = \mathbb{E}[L_i(Y_i, W_n; \theta)]$$

be the population log-likelihood of the game model with n individuals, where the second equality comes from the implicitly assumption that all individuals are created equal on the network before its realization. Given the identification and regularity conditions, θ_0 maximizes $L(\cdot; n)$ for every n by a standard argument for MLE.

Following Xu (2011), the MLE is defined as follows:

$$\hat{\theta}_{MLE} = \arg \max_{\theta \in \Theta} L_n(\theta)$$

where $L_n(\theta) = \frac{1}{n} \sum_{i=1}^n L_i(Y_i, W_n; \theta)$. The difference between $L(\theta; n)$ and $L_n(\theta)$ is essential: $L(\theta; n)$ is a deterministic function of θ that represents population log-likelihood of our n -individual network game, while $L_n(\theta)$ is a random function of θ that is a sample analog of $L(\theta; n)$. Next lemma provides consistency and asymptotic normality for $\hat{\theta}_{MLE}$.

Assumption 6. *There exists a non-singular $(d + d_s) \times (d + d_s)$ matrix I_{θ_0} , such that*

$$\mathbb{E} \left[\frac{\partial}{\partial \theta} L_i(Y_i, W_n; \theta) \times \frac{\partial}{\partial \theta'} L_i(Y_i, W_n; \theta) \right] \rightarrow I_{\theta_0}$$

Assumption 7. \mathcal{S}_X is bounded and Θ is compact.

Assumption 8. θ_0 is an interior point of Θ .

The matrix I_{θ_0} is the Fisher information when the number of players goes to infinity. The non-singularity condition of I_{θ_0} as well as assumptions 7 and 8 are standard in the MLE literature.

Lemma 2. Under conditions assumptions 1, 2 and 5 to 7, we have $\hat{\theta}_{MLE} \xrightarrow{p} \theta_0$.

Moreover, suppose assumption 8 holds. Then

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta_0) \xrightarrow{d} N(0, I_{\theta_0}^{-1})$$

Proof. The proof obtains by a similar argument in Xu (2011), and hence is omitted here.

Though the most efficient, the MLE is extremely difficult to compute in a large network: for each iteration under a parameter value θ , we need to solve the fixed point in an n -individual game where n is large. In this paper, we suggest a different estimation procedure, called as “nested pseudo likelihood” (NPL), which is simple to implement and provides standard error of the estimator directly. The proposed estimation method is attractive from a practitioner’s perspective.

In a seminal paper, Aguirregabiria and Mira (2007) introduced NPLE for estimating dynamic discrete games that also involve computational difficulties of obtaining an equilibrium in a Bayesian game with infinite time horizon. For the same reason, by using the nested pseudo likelihood algorithm we can avoid to compute fixed points in a large vector space.

2.5.2. Nested Pseudo Likelihood Estimation

To define NPLE, we first introduce some notation. Let

$$q_i(Y_i, W_n; \theta, P_n) = Y_i \ln \Gamma_i(P_n, W_n; \theta) + (1 - Y_i) \ln(1 - \Gamma_i(P_n, W_n; \theta))$$

be the pseudo log-likelihood of individual i 's choice (conditional on W_n). In the definition of q_i , Γ_i is not the equilibrium choice probability, but the best response function. Moreover P_n is a generic choice probability profile instead of the actual equilibrium choice probability profile. In this sense, q_i is a pseudo log-likelihood function. Let further

$$Q_n(\theta, P_n) = \frac{1}{n} \sum_{i=1}^n q_i(Y_i, W_n; \theta, P_n)$$

When $P_n = P_n^*(W_n; \theta_0)$, $Q_n(\cdot, P_n^*(W_n; \theta_0))$ becomes a true likelihood function of θ , which is maximized at θ_0 . It should also be noted that $L_n(\theta) = Q_n(\theta, P_n^*(W_n; \theta))$.

Following Aguirregabiria and Mira (2007), we now define the NPLE using an interactive algorithm, which begins with an arbitrary initial guess $P_n^{(0)}$ for $P_n^*(W_n; \theta_0)$. Note that $P_n^{(0)}$ need not be a consistent estimator of $P_n^*(W_n; \theta_0)$. For example, we could take $P_n^{(0)} = (0, \dots, 0)'$. Formally, the NPL algorithm is described as follows:

Step 1: Let

$$\hat{\theta}_{(1)} = \arg \max_{\theta \in \Theta} Q_n(\theta, P_n^{(0)}),$$

$$P_n^{(1)} = \Gamma(P_n^{(0)}, W_n; \hat{\theta}_{(1)})$$

Step 2: This step is an iterative procedure: for $k \geq 2$, let

$$\hat{\theta}_{(k)} = \arg \max_{\theta \in \Theta} Q_n(\theta, P_n^{(k-1)}),$$

$$P_n^{(k)} = \Gamma(P_n^{(k-1)}, W_n; \hat{\theta}_{(k)})$$

This procedure stops at the K -th iteration when $\|\hat{\theta}_{(K)} - \hat{\theta}_{(K-1)}\|$ is less than a predetermined tolerance, e.g. 10^{-6} .

Step 3: Let $\hat{\theta}_{(K)}$ be our NPLE, i.e. $\hat{\theta}_{NPL} = \hat{\theta}_{(K)}$. Moreover, let $\hat{P}_n = P_n^{(K)}$.

As also pointed by Aguirregabiria and Mira (2007), in general it is unclear whether the NPL algorithm converges or not. Therefore, we investigate this issue by using Monte Carlo experiments, in which the NPL algorithm converges well. By definition, our NPLE is essentially a fixed point estimator that we obtain using the proposed iterative algorithm if it converges. Specifically, the fixed point is defined in the following equation system in $(\hat{\theta}_{NPL}, \hat{P}_n)$:

$$\hat{\theta}_{NPL} = \arg \max_{\theta \in \Theta} Q_n(\theta, \hat{P}_n) \text{ and } \hat{P}_n = \Gamma(\hat{P}_n, W_n; \hat{\theta}_{NPL}) \quad (7)$$

See Aguirregabiria and Mira (2007) for more details. The motivation for NPLE comes from the following observation: let $Q(\theta, P_n; n) = \mathbb{E}Q_n(\theta, P_n)$ be the population objective function. Because $Q(\cdot, P_n^*(W_n; \theta_0); n)$ is an actual likelihood function of θ in which $P_n^*(W_n; \theta_0)$ is the underlying choice probability profile derived from equilibrium conditions, then θ_0 maximizes $Q(\cdot, P_n^*(W_n; \theta_0); n)$ for all n , i.e.

$$\theta_0 = \arg \max_{\theta \in \Theta} Q(\theta, P_n; n) \text{ and } P_n = \Gamma(P_n, W_n; \theta_0) \quad (8)$$

The above discussion is summarized by the following lemma.

Lemma 3. *Suppose regular conditions in the MLE hold. For any n , θ_0 solves the following equation*

$$\theta = \arg \max_{\theta' \in \Theta} Q(\theta', P_n^*(W_n; \theta); n) \quad (9)$$

Proof. See Appendix B.2.

2.5.3. Consistency

To establish the consistency, we need to make an additional assumption. Note that equation (9) might admits multiple solutions, which means that θ_0 is not identified by the pseudo likelihood function. For this reason, we simply rule out such a possibility by the following assumption.

Assumption 9. θ_0 uniquely solves the following equation

$$\theta = \lim_{n \rightarrow \infty} \arg \max_{\theta' \in \Theta} Q(\theta', P_n^*(W_n; \theta); n) \quad (10)$$

Assumption 9 implicitly requires that the population objective function $Q(\cdot, P_n^*(W_n; \theta); n)$ converges to some limit function of θ' . The limit is taken over the number of individuals in the network which is the sample size. The existence of such a limit is due to the fact that as n increases, the equilibrium choice probability of any given individual i will converges to its limit. This is because the additional individuals will be located farther away from i in the network under our assumption 2. See Appendix A.2 for a discussion on the network stability condition. Assumption 9 can be replaced by a weaker assumption that θ_0 is a unique solution in an open ball around it, see Aguirregabiria and Mira (2007) for more details. Hiroyuki and Shimotsu (2012) discuss a local condition under which the NPL estimator is consistent. The condition is a local contraction property and they also propose a modified equilibrium update rule to obtain a better contraction property and more consistent NPL estimator.

Given the conditions specified above, we show that the NPLE is consistent.

Theorem 3. Suppose assumptions 1, 2, 4, 5, 7 and 9 hold. Then $\hat{\theta}_{NPL} \xrightarrow{p} \theta_0$.

Proof. See Appendix B.3.

2.5.4. Asymptotic Normality

Here we derive the limiting distribution for our NPLE. For the illustration purpose, we also investigate the asymptotic properties of the ML estimator as well as the pseudo maximum likelihood estimator (PMLE). Following Aguirregabiria and Mira (2007), the PMLE is defined as follows $\hat{\theta}_{\text{PMLE}} = \arg \max_{\theta \in \Theta} Q_n(\theta, P_n^*(W_n; \theta_0))$. Note that PMLE is an infeasible estimator since $P_n^*(W_n; \theta_0)$ is unknown while MLE is not practical due to its heavy computational burden. To proceed, we first examine the FOCs for these three estimators.

$$\text{NPLE: } \frac{\partial Q_n(\hat{\theta}_{\text{NPLE}}, P_n^*(W_n; \hat{\theta}_{\text{NPLE}}))}{\partial \theta} = 0$$

$$\text{PMLE: } \frac{\partial Q_n(\hat{\theta}_{\text{PMLE}}, P_n^*(W_n; \theta_0))}{\partial \theta} = 0$$

$$\text{MLE: } \frac{\partial Q_n(\hat{\theta}_{\text{MLE}}, P_n^*(W_n; \hat{\theta}_{\text{MLE}}))}{\partial \theta} + \frac{\partial Q_n(\hat{\theta}_{\text{MLE}}, P_n^*(W_n; \hat{\theta}_{\text{MLE}}))}{\partial P_n} \cdot \frac{\partial P_n^*(W_n; \hat{\theta}_{\text{MLE}})}{\partial \theta} = 0$$

Clearly, the differences are essential: compared with the PMLE, the NPLE introduces an additional bias through the choice probability profile $P_n^*(W_n; \hat{\theta}_{\text{NPLE}})$. The MLE differs from the NPLE in the sense that θ affects equilibrium choice probabilities of friends. Moreover, these three FOCs reflect the different amount of information used in these likelihood approaches. Specifically, the PMLE exploits information contained in the partial equilibrium by using the best response equations on the equilibrium path while the MLE is indeed a general equilibrium approach. Further, we view our NPLE as

a method in between. Next, we establish the asymptotic normality for our NPL estimator.

To begin with, we make the following assumption.

Assumption 10. *There exist non-singular $(d + d_s) \times (d + d_s)$ matrices $V_1(\theta_0)$ and $V_2(\theta_0)$ such that*

$$-\mathbb{E} \left[\frac{\partial^2 Q_n(\theta_0, P_n^*(W_n; \theta_0))}{\partial \theta \partial \theta'} \middle| W_n \right] \xrightarrow{p} V_1(\theta_0)$$

$$\mathbb{E} \left[\frac{\partial^2 Q_n(\theta_0, P_n^*(W_n; \theta_0))}{\partial \theta \partial P_n} \cdot \frac{\partial P_n^*(W_n; \theta_0)}{\partial \theta'} \middle| W_n \right] \xrightarrow{p} V_2(\theta_0)$$

Moreover, $V(\theta_0) = V_1(\theta_0) - V_2(\theta_0)$ is positive definite.

Assumption 10 is a high level condition which imposes restrictions on the sequence of networks (indexed by the number of individuals) in the asymptotic analysis. Such a condition could be derived from primitive assumptions on W_n , e.g. X_i are i.i.d. and the distribution of the number of friends NF_i converges to some limiting distribution as n goes to infinity. In our setting, we can express the conditions in assumption 10 in a more explicit manner. To illustrate, we first introduce some notation. For $s, s' \in \{L, H\}$, let $Z_{\alpha i s s'}^* = Z_{\alpha i s'} \cdot 1\{S_i = s\}$, $Z_{\gamma i s s'}^* = Z_{\gamma i s} \cdot 1\{S_i = s\}$. Let further $X_{\alpha i}^* = (Z_{\alpha i L L}^*, Z_{\alpha i L H}^*, Z_{\alpha i H L}^*, Z_{\alpha i H H}^*)$, $X_{\beta i}^* = X_i'$, $X_{\gamma i}^* = (Z_{\gamma i L L}^*, Z_{\gamma i L H}^*, Z_{\gamma i H L}^*, Z_{\gamma i H H}^*)$, and $X_i^* = (X_{\alpha i}^*, X_{\beta i}^*, X_{\gamma i}^*)'$. First, note that

$$\mathbb{E} \left[\frac{\partial^2 Q_n(\theta_0, P_n^*(W_n; \theta_0))}{\partial \theta \partial \theta'} \middle| W_n \right] = -n \text{Var} \left(\frac{\partial Q_n(\theta_0, P_n^*(W_n; \theta_0))}{\partial \theta} \middle| W_n \right)$$

where the first equality comes from the conditional independence, the second step comes from information equality and the last equality comes use zero conditional mean.

Furthermore, with simple calculation, we have

$$\frac{\partial Q_n(\theta_0, P_n^*(W_n; \theta_0))}{\partial \theta} = \frac{1}{n} \sum_{i=1}^n X_i^* (Y_i - p_i^*(W_n; \theta_0))$$

Therefore

$$nVar\left(\frac{\partial Q_n(\theta_0, P_n^*(W_n; \theta_0))}{\partial \theta} \middle| W_n\right) = \frac{1}{n} \sum_{i=1}^n X_i^* X_i^{*'} p_i^*(W_n; \theta_0) (1 - p_i^*(W_n; \theta_0))$$

Theorem 4. Suppose assumptions 1, 2, 5 and 7 to 10 hold, we have

$$\sqrt{n}(\hat{\theta}_{NPL} - \theta_0) \xrightarrow{d} N(0, \Omega(\theta_0)),$$

$$\Omega(\theta_0) = V^{-1}(\theta_0) \cdot V_1(\theta_0) \cdot V^{-1}(\theta_0)$$

Proof. See Appendix B.4.

2.5.5. Simulation-Based Estimation of Variance-Covariance Matrix

In this subsection, we provide a simulation-based method to consistently estimate the first derivative of choice probabilities respect to the parameter, $\frac{\partial P_n^*(W_n; \theta_0)}{\partial \theta}$. After we obtain our NPL estimator $\hat{\theta}_{NPL}$, we draw a sequence of parameters from a small open ball of $\hat{\theta}_{NPL}$ and denote them as $\{\theta_i\}_{i=1}^T$ by $\theta_i = \hat{\theta}_{NPL} + \varepsilon_i$ where ε_i is drawn from a uniform distribution on $[-a, a]$. With each θ_i , we compute the choice probabilities through best response function, i.e., $S_n^{(i)} = \Gamma(P_n^*, W_n; \theta_i)$. Therefore we have

$$S_n^{(i)} - P_n^* = \frac{\partial P_n^*(W_n; \hat{\theta}_{NPL})}{\partial \theta} (\theta_i - \hat{\theta}_{NPL}) + \xi_i$$

Therefore we have our estimator of the variance-covariance matrix:

$$\frac{\partial \widehat{P}_n^*(W_n; \hat{\theta}_{NPL})}{\partial \theta} = \left[\frac{1}{T} \sum_{i=1}^T S_n^{(i)} (\theta_i - \hat{P}_{NPL}) \right] \left[\frac{1}{T} \sum_{i=1}^T (\theta_i - \hat{P}_{NPL}) (\theta_i - \hat{P}_{NPL})' \right]^{-1}$$

With consistency of $\hat{\theta}_{NPL}$ and arbitrarily small a , we can easily show that $\frac{\partial \hat{P}_n^*(W_n; \hat{\theta}_{NPL})}{\partial \theta}$ is a consistent estimator for $\frac{\partial P_n^*(W_n; \theta_0)}{\partial \theta}$. Therefore we can obtain consistent estimator for Ω through such a simulation-based method.

2.5.6. Testing Hierarchy Effects

It is interesting to look at the asymmetry in social interactions which captures the hierarchy effects. Intuitively, we expect that high status friends would have larger peer effects. Through all results established above, we are able to quantify the hierarchy effects of how large is the gap between peer effects from friends of different status. We formally write out the null hypothesis as follows

$$H_0: \alpha(y, s, H) - \alpha(y, s, L) = 0, y = 0, 1; s = L, H$$

Based on Theorem 4, we can construct standard Wald-style test for above null hypotheses. With consistent estimators for both parameters and their standard error, we can detect the size of the hierarchy effects and how significant they are. This is important for policy analysis on how to targeting and reconstructing the network structure on some goals, e.g, fairness.

2.6. MONTE CARLO EXPERIMENTS

In this section we present evidence for the performance of the NPL estimators in a simple school network on college attendance decision. For student i , the utility from attending college is $X_i' \beta + \sum_{j \in F_i} \mathbb{E}(\alpha(Y_j, S_i, S_j)) + \epsilon_i$ where Y_j denotes the college attendance decision of student j . The utility of not attending college is normalized to zero. Each student i is associated with a social status, H or L , decided by number of friend nominations from other students. The true parameter of exogenous effect is invariant

across the different Monte Carlo experiments, i.e. $\beta_0 = -4$. To illustrate how our NPL estimator performs with different hierarchy effects, we execute three experiments with true interactions parameters set as follows

Monte Carlo I: $\{\alpha_0 = (0.1, 0.2, 0.3, 0.4) \text{ and } \gamma_0 = (0.1, 0.2, 0.3, 0.4)\}$

Monte Carlo II: $\{\alpha_0 = (0.1, 0.1, 0.1, 0.1) \text{ and } \gamma_0 = (0.1, 0.2, 0.3, 0.4)\}$

Monte Carlo I: $\{\alpha_0 = (0.1, 0.1, 0.1, 0.1) \text{ and } \gamma_0 = (0.1, 0.1, 0.1, 0.1)\}$

Experiment I has both heterogenous peer effects and hierarchy effects. Experiment II imposes homogenous hierarchy effects while Experiment III take peer effects as homoeogenous. X_i is drawn from standard uniform distribution and ϵ_i conforms to standard Logistic distribution. With randomly drawn X_i, ϵ_i and the true parameter, we solve the unique equilibrium, P_n^* . Then we calculate Y_i through

$$Y_i = 1 \left\{ X_i' \beta + \sum_{j \in F_i} [\alpha(0, S_i, S_j) + \gamma(S_i, S_j) \cdot p_j^* + \epsilon_i] > 0 \right\}$$

where $Y_i = 1$ denotes “attend college” and $Y_i = 0$ otherwise. By construction, the conditional independence, unique equilibrium, and rank conditions are all satisfied by these designs. From our asymptotic results of $\hat{\theta}_{NPL}$, we know that $MSE = O_p\left(\frac{1}{n}\right)$ and $SD = O_p\left(\frac{1}{\sqrt{n}}\right)$. Therefore, with $\{(Y_i, X_i, S_i, F_i)\}_{i=1}^n$, we implement our nested pseudo likelihood algorithm to estimate the parameter of interest. For each experiment design, we simulate $NR = 1000$ samples and calculate summary statistics from empirical distribution of estimators $\hat{\theta}_{NPL}$ from these samples. We focus on three statistics: mean, standard deviation (SD) and mean squared error (MSE). MSE is estimated using the

estimators and the knowledge of our true parameters in the setting. In each Monte Carlo experiment, we consider three different number of individuals, $n = 1200, 2400, 4800$ to investigate at the orders of MSE and standard deviation. Table 1 reports performance of $\hat{\theta}_{NPL}$ in the design with heterogenous peer effects and hierarchy effects in terms of mean and standard deviation. For all three sample sizes, the means of $\hat{\theta}_{NPL}$ are quite close to the true parameter which confirm the consistency. The MSE in Table 2 is halved when the sample size double which is in accordance with the order from the asymptotic properties, i.e. $MSE = O_p\left(\frac{1}{n}\right)$. The standard errors are with reasonable size for discrete choice model. We compare our results with that of Aguirregabiria and Mira (2007) which execute the Monte Carlo simulations with size sample equal 2000. Our results with $n = 2400$ quite similar as that of Aguirregabiria and Mira (2007) in terms of standard error. Aguirregabiria and Mira (2007) consider the oligopolistic competition and the validity of NPL estimation comes from the independent market assumption. In our social network scenario, this independence condition is violated; however, we recover the validity from the incompleteness of the social network and the constraint on the number of friends.

Table 3 reports performance of $\hat{\theta}_{NPL}$ in a setting that homogenous peer effects from null action, $y = 0$ and heterogenous peer effects from $y = 1$. The summary statistics are similar as those from fully heterogenous setting in experiment I: the mean is close to the true parameter, the MSE in Table 4 is with order $1/n$ and the standard error diminishes in a rate of $1/\sqrt{n}$. In experiment III, we have that all peer effects are the same

and focus on the hierarchy effects. The results turn out to be as good as those from experiment I and II. We present those summary statistics in Tables 5 and 6.

Table 1 Mean and Standard Deviation I

n	$\hat{\beta}$	$\hat{\alpha}$				$\hat{\gamma}$			
1200	-4.045	0.097	0.209	0.297	0.412	0.107	0.194	0.312	0.391
	<i>-0.278*</i>	<i>-0.187</i>	<i>-0.128</i>	<i>-0.180</i>	<i>-0.172</i>	<i>-0.467</i>	<i>-0.254</i>	<i>-0.477</i>	<i>-0.286</i>
2400	-4.023	0.093	0.198	0.300	0.402	0.112	0.211	0.308	0.403
	<i>-0.196</i>	<i>-0.139</i>	<i>-0.087</i>	<i>-0.126</i>	<i>-0.125</i>	<i>-0.348</i>	<i>-0.180</i>	<i>-0.334</i>	<i>-0.208</i>
4800	-4.010	0.099	0.202	0.300	0.401	0.104	0.196	0.308	0.400
	<i>-0.144</i>	<i>-0.096</i>	<i>-0.060</i>	<i>-0.085</i>	<i>-0.089</i>	<i>-0.242</i>	<i>-0.120</i>	<i>-0.224</i>	<i>-0.154</i>

*Standard Errors are denoted with italic font

Table 2 Mean Square Error I

n	$\hat{\beta}$	$\hat{\alpha}$				$\hat{\gamma}$			
1200	0.079	0.035	0.017	0.033	0.030	0.218	0.064	0.228	0.082
2400	0.039	0.019	0.008	0.016	0.016	0.121	0.033	0.112	0.043
4800	0.021	0.009	0.004	0.007	0.008	0.058	0.014	0.050	0.024

Table 3 Mean and Standard Deviation II

n	$\hat{\beta}$	$\hat{\alpha}$				$\hat{\gamma}$			
1200	-4.048	0.084	0.105	0.103	0.100	0.134	0.197	0.297	0.414
	<i>-0.309</i>	<i>-0.193</i>	<i>-0.102</i>	<i>-0.171</i>	<i>-0.143</i>	<i>-0.602</i>	<i>-0.311</i>	<i>-0.536</i>	<i>-0.354</i>
2400	-4.022	0.104	0.101	0.101	0.103	0.086	0.200	0.305	0.392
	<i>-0.213</i>	<i>-0.135</i>	<i>-0.073</i>	<i>-0.121</i>	<i>-0.102</i>	<i>-0.418</i>	<i>-0.222</i>	<i>-0.382</i>	<i>-0.253</i>
4800	-4.012	0.101	0.100	0.097	0.104	0.103	0.205	0.304	0.399
	<i>-0.152</i>	<i>-0.090</i>	<i>-0.051</i>	<i>-0.081</i>	<i>-0.070</i>	<i>-0.291</i>	<i>-0.157</i>	<i>-0.256</i>	<i>-0.176</i>

Table 4 Mean Square Error II

n	$\hat{\beta}$	$\hat{\alpha}$				$\hat{\gamma}$			
1200	0.098	0.037	0.011	0.029	0.020	0.363	0.097	0.287	0.126
2400	0.046	0.018	0.005	0.015	0.010	0.175	0.049	0.146	0.064
4800	0.023	0.008	0.003	0.007	0.005	0.085	0.025	0.066	0.031

Table 5 Mean and Standard Deviation III

n	$\hat{\beta}$	$\hat{\alpha}$				$\hat{\gamma}$			
1200	-4.021	0.098	0.205	0.304	0.399	0.107	0.092	0.097	0.106
	-0.276	-0.186	-0.118	-0.162	-0.155	-0.504	-0.274	-0.460	-0.313
2400	-4.021	0.098	0.200	0.299	0.400	0.103	0.107	0.111	0.108
	-0.201	-0.133	-0.084	-0.120	-0.114	-0.351	-0.197	-0.336	-0.222
4800	-4.014	0.101	0.203	0.298	0.403	0.101	0.098	0.107	0.101
	-0.149	-0.091	-0.057	-0.083	-0.081	-0.261	-0.134	-0.233	-0.154

Table 6 Mean Square Error III

n	$\hat{\beta}$	$\hat{\alpha}$				$\hat{\gamma}$			
1200	0.077	0.035	0.014	0.026	0.024	0.254	0.075	0.212	0.098
2400	0.041	0.018	0.007	0.014	0.013	0.123	0.039	0.113	0.049
4800	0.022	0.008	0.003	0.007	0.007	0.068	0.018	0.054	0.024

2.7. COLLEGE ATTENDANCE

In this section, we apply our model to study the college attendance decisions among high school student using Add Health dataset. Our results show that there are significant hierarchy effects in the peer effects of college attendance choices. Besides peer effects, the gpa is a key characteristics that affects the education decision.

2.7.1. Add Health Dataset

The National Longitudinal Study of Adolescent Health (Add Health) is a longitudinal study of a nationally representative sample of adolescents in grades 7-12 in the United States during the 1994-95 school year. After the first wave, there are three following wave survey, keeping track on the adolescents which provides the college attendance data. Add Health combines longitudinal survey data on respondents' social and economic features with contextual data on the family, friendships and peer groups. Therefore, this dataset provides unique opportunity to study peer effects and hierarchy effects. We are interested in the peer effects in decisions of whether continuing college education after graduation from high school. Whether there are asymmetric peer effects in students' choice is interesting and meaningful for policy analysis. To be more specific, we are interested in the monotonicity of peer effect, i.e., the decisions of students with high status have larger impact on the choice of students in the low-status group. This monotonicity has policy implication that government could target on the high-status group to increase the whole enrollment rate of high school students. In the Add Health dataset, each student has nominations of at most five male friends and at most five female friends. This structure provides a social network with direct links. Though students have at most 10 out-links, they may have much more in-links. For instance, a student is attractive and nominated by 20 students as their friends, then this student has 20 in-link. The in-link structure determines the hierarchy among students. We use demographic characteristics similar as those in the literature, see Dynarsky (2003); Linsenmeier et al. (2006); Garibaldi et al. (2012), such as age, household income, GPA,

parents' education, race information, gender, etc. Table 7 summarizes the statistics of the college attendance variable and demographic variables. There are 4678 observations in the final analysis sample. The college enrollment rate was 57%. Though good dataset for peer effects study, Add Health dataset suffers from the same problem of survey data, e.g. for parents' education, many respondents report 0 year education. For household income, respondents were likely to report much lower values. The main missing data issue for our analysis is the reluctance of nominations for best friends. Due to missing friendship, we combine four years high school students for our college attendance analysis. Our results may be biased due to the missing information of friendship and the not static setting of four years enrollment data.

Table 7 Summary of Statistics

Variable	Mean	Std. Dev
Age	16.530	1.560
Female	0.540	0.500
Household Income	7.990	3.040
Mother Education	0.660	2.010
Father Education	1.910	3.180
Overall GPA	2.510	0.800
American Indian	0.040	0.190
Asian	0.060	0.240
Black	0.150	0.360
Hispanic	0.160	0.370
White	0.720	0.450
College Attendance	0.570	0.490

2.7.2. Peer Effects in College Attendance

We can easily imagine that when high school students make college attendance decision, the choices of their friends deliver impacts. Previous studies on college attendance investigate the effects of demographic characteristics as well as contextual variables, e.g. whether parents attended college, whether single-parent household, family size, female or male, family income, high school grade, whether minority, whether student aid, etc.(Manski and Wise (1983); Dynarsky (2003); Linsenmeier et al. (2006) to new only a few). However, the peer effects are not thoroughly studies in college attendance decision. To the best of our knowledge, we are the first one explicitly dealing with college attendance with comprehensive peer effects analysis. Table 8 reports the estimation results. We find that the overall GPA is the key factor that determines the college enrollment. This indicates that some high school students did not attend college because they did not receive any offer due to low GPA.

The white students are less likely to attend college for advanced education. This may be due to the good outside options for white students. Female students are slightly less likely to attend college from our result which provides evidence of gender discrimination in the 90's. An interesting finding from our result is that household income has small impact on college attendance. This provides evidence that students confronted less financial constraints. We also find that the older, the less probabilities to attend college which coincides with our intuition. The most important finding in our empirical study of college attendance is that there are strong hierarchy effects in peer effects. The peer effects from friends' not attending college ($Y_j = 0$) are all negative and

the high status friends have much larger impact than low status students for the non-action, i.e.

$$\alpha(0, L, H) = -1.038$$

$$\alpha(0, L, L) = -0.296,$$

but similar effect from attending decision

$$\alpha(1, L, H) = 0.415$$

$$\alpha(1, L, L) = 0.403.$$

We also find significant asymmetric peer effects that high status friends have stronger peer effects on low status students than vice versa, i.e.

$$\alpha(0, L, H) = -1.038$$

$$\alpha(0, H, L) = -0.120.$$

Also, we have

$$\alpha(1, L, H) = 0.415$$

$$\alpha(1, H, L) = 0.276.$$

To summarize, we find strong hierarchy effects that high status students delivers larger peer effects. Further, not attending college has stronger peer effects which support the strong peer effects on the negative side. Similar results are found in Carrell et al. (2013).

Table 8 Estimation Results

Variable	Estimator	Std. Dev	P-value
Age	-0.186	0.010	0.000
Female	-0.221	0.070	0.001
Household Income	0.095	0.012	0.000
Mother Education	-0.017	0.016	0.311
Father Education	-0.013	0.011	0.213
Overall	1.312	0.052	0.000
American Indian	-0.235	0.174	0.176
Asian	0.026	0.175	0.883
Black	-0.045	0.140	0.747
Hispanic	-0.122	0.105	0.246
White	-0.532	0.119	0.000
Peer Effects	Estimator	Std. Dev	P-Value
$\alpha(0, L, L)$	-0.296	0.298	0.322
$\alpha(0, L, H)$	-1.038	0.325	0.001
$\alpha(0, H, L)$	-0.120	0.498	0.809
$\alpha(0, H, H)^*$			
$\gamma(L, L)$	0.699	0.478	0.144
$\gamma(L, H)$	1.453	0.517	0.005
$\gamma(H, L)$	0.396	0.836	0.636
$\gamma(H, H)^*$			

*: Seldom “High”-“High” Pair, therefore no enough variation

2.8. CONCLUSION

We provide theoretical and empirical analyses of hierarchy effects and peer effects in the large network game. We contribute to the literature by incorporating social status in peer effects and providing an easy implemented estimation method to recover the parameter of interest.

CHAPTER III

SEMIPARAMETRIC PANEL DATA TRUNCATED REGRESSION MODEL WITH
FIXED EFFECTS, WITH CHUNRONG AI AND HONGJUN LI

3.1 INTRODUCTION

Truncated model is widely used to deal with incomplete observations of the dependent variable, which are commonly encountered in empirical economic analysis. For example, in labor economics, the wages and other demographic characteristics of unemployed people are not observed and therefore those unemployed are truncated from the sample. If we are interested in a relationship between the income and the education, we have to take into account the missing data issue. Econometricians suggest several approaches to solve this missing data problem. Main streams include truncated regression models and partial identified models. Truncated regression model is one of the Tobit models which deal with data with some part totally missed, both dependent variable and independent variables. With latent variables, truncated regression models propose consistent estimators for the parameters and give the corresponding asymptotic properties (i.e. Powell (1986), Honoré (1992), Honoré and Powell (1992), Lewbel and Linton (2002), etc.). Partial identification models relax the point identified estimation and allow unknown properties for the truncated sample. Researchers achieve bound/interval estimation and allow more general structure of the missing data in recent years (i.e. Manski (1995, 2003, 2005), Horowitz and Manski (1995, 2006), Molinari

(2008), etc.) In this paper, we focus on the truncated regression model and contribute to the literature by allowing more flexible regression form.

We consider a partially linear panel data truncated regression model with fixed effects. Economic theories seldom suggest linear relationships between economic variables and there are several applications in which parametric models do not fit the real data well (i.e. Härdle (1990), Horowitz and Lee (2002), Perrigne and Vuong (2011), etc.). Thus more flexible models (Semiparametric and Nonparametric) are desirable when we are not confident in making correct model specification. Nonparametric models receive much attention from the researchers in the last three decades (See Li and Racine (2007), Pagan and Ullah (1999)). However, nonparametric estimations suffer from three main drawbacks: (1) Curse of dimensionality; (2) The results are hard to display, communicate and interpret; (3) Fail to allow extrapolation (See Horowitz (2009) for details). Unlike nonparametric models, semiparametric models have the convenience of parametric structure and the flexibility of nonparametric form and avoid the drawbacks of both. Our main contribution in this paper is applying sieve method to the nonparametric components of the nonstandard panel data truncated regression model with fixed effects while keeping the \sqrt{n} convergent rate of the parametric estimator.

The identification and estimation of truncated regression model are quite different from the standard linear regression models even in the cross-sectional case. The conventional OLS method is not consistent since the distribution from the truncated sample is rescaled so that the conditional independence of the regressors and error do not hold. For the truncated regression model, the observations are drawn from the population

by conditioning on certain event. Specific definition of truncated data will be given below. To achieve consistent estimation, Powell (1986) constructs a trimmed least squares estimator for cross sectional truncated regression model. With symmetrical trimming technique, Powell recovers the symmetry of the distribution of the dependent variable and gets conditional moment restriction at the cost of dropping part of the data. For panel data, the story is different since there are fixed effects or random effects. Panel data models gains much more attention in recent years. While random effects presents in some experimental data, researchers usually consider the one dimension effects to be fixed. Time invariant fixed effects terms are used to capture the endogeneity. We can use differencing method to get rid of the fixed effects terms can without directly estimate them. It is well known that the maximum likelihood estimators with direct estimation of fixed effects terms are consistent. However, for a truncated regression model with fixed effects, due to the nonlinear nature of these models, we cannot get rid of the fixed effects by simple differencing. Honoré (1992) extends the idea of symmetrical trimming of Powell (1986) to panel data model with fixed effects and proposes a trimming LAD estimation method for panel data truncated regression model with fixed effects. He establishes consistency and \sqrt{n} asymptotic normality of his proposed estimator. Our model is a direct extension of Honoré's model by incorporating nonparametric component. While similar model, our identification strategy and estimation method are different. Our semiparametric model has the advantages to deliver economic interpretation from the parametric part and conduct policy analysis by extrapolation while keeping the flexibility.

For identification, we derive the conditional moment restrictions following similar arguments in Honoré (1992). The difference is that our new conditional moment restriction involves the nonparametric part due to the partially linear form in the regression. For estimation, we apply a sieve estimation method. Sieve method is popular due to its \sqrt{n} convergent rate in asymptotic sense and easy implementation. For the convergence rates of sieve methods, we refer to Andrews (1991) and Newey (1995, 1997), Chen (2007). Especially for those models containing nonsmooth objective function in parameters and unknown functions. In our model, the objective function is nonstandard, nonsmooth in the parametric parameters and nonparametric unknown function. With sieve estimation methods we can directly plug in series approximation in the objective function and solve the series coefficients easily while not affecting the linear parameter and estimated variance. We establish the consistency and \sqrt{n} asymptotic normal distribution of the estimator for the linear parameter. The conditions we provide in this paper are primitive and easy to check.

The rest of the paper is organized as follows. Section 2 gives an introduction of panel data truncated regression models with fixed effects and the idea of symmetrical trimming and propose our estimator. Section 3 shows our identification strategy. Section 4 derives the consistency for the estimator and section 5 establishes the asymptotic normality of our estimator. Finally, section 6 concludes the paper.

3.2 PANEL DATA TRUNCATED REGRESSION MODEL WITH FIXED EFFECTS

We consider the following semiparametric panel data truncated regression model:

$$Y_{it}^* = \alpha_i + X_{it}'\beta + h(Z_{it}) + \epsilon_{it} \quad (1)$$

where Y_{it}^* is a latent variable, X_{it} is a d_1 -dimension vector of explanatory variables, Z_{it} is a d_2 -dimension vector of contextual variables and ϵ_{it} is the error terms. β is a vector of parameters of primary interest with dimension d_1 , α_i is the fixed effects and $h(\cdot)$ is an unknown function. We assume $\beta \in \mathcal{B}$, where \mathcal{B} is a compact subset of \mathbb{R}^{d_1} , and $h \in \mathcal{H}$, where \mathcal{H} is an infinite dimensional subset of a Banach space with norm $\|\cdot\|_s$, such as the space of bounded continuous functions with the sup-norm $\|h\|_s = \sup_z |h(z)|$, or the space of square integrable functions with the root mean squared norm $\|h\|_s = \sqrt{\mathbb{E}[h^2(Z)]}$. We denote $\theta \equiv (\beta, h)$.

Typical panel data have a large number of individuals and few time periods. So we assume the number of time periods is fixed. Without loss of generality, we assume the number of time periods is two. We can easily extend our results to the cases of $T > 2$. For the Truncated regression model, the observations $\{(Y_{it}, X_{it}, Z_{it}): t = 1, 2; i = 1, \dots, n\}$ are a random sample from the distribution of $\{(Y_{it}^*, X_{it}, Z_{it}): t = 1, 2; i = 1, \dots, n\}$ induced by conditioning on the event $\{Y_{it}^* > 0\}$. That is, when $Y_{it}^* \leq 0$, we can observe neither Y_{it}^* nor the corresponding X_{it} and Z_{it} . We use set of $k(n)$ sieve basis functions to approximate h_0 , i.e. $h_{k(n)}(Z_i) \equiv P_{k(n)}(Z_i)' \delta_{k(n)}$. Denote the space of the series approximation function as $\mathcal{H}_{k(n)}$. $P_{k(n)}(Z_i)$ could be power series, Fourier series, splines, wavelets and many other series or combined ones. To achieve regular

asymptotic properties, $k(n)$ is set to satisfy: $k(n) \rightarrow \infty$ as $n \rightarrow \infty$ and $k(n)/n \rightarrow 0$ as $n \rightarrow \infty$.

Assumption 1. *The conditional distribution of $(\varepsilon_{i1}, \varepsilon_{i2})$ given $(\alpha, X_{i1}, X_{i2}, Z_{i1}, Z_{i2})$ is continuous with finite density for all i .*

Assumption 2. *The distribution of $\varepsilon_{i1} - \varepsilon_{i2}$ conditional on $\varepsilon_{i1} + \varepsilon_{i2}$ and on $(\alpha, X_{i1}, X_{i2}, Z_{i1}, Z_{i2})$ is strictly unimodal and symmetric around 0.*

With assumption 1 and 2, and applying the symmetrical trimming approach in Honoré (1992) we construct our estimator as

$$\hat{\theta}_{k(n)} = \arg \max_{\beta \in \mathcal{B}, h_{k(n)} \in \mathcal{H}_{k(n)}} R_n(\theta_{k(n)})$$

where

$$\begin{aligned} R_n(\theta_{k(n)}) = & -\frac{1}{n} \sum_{i=1}^n \left[\left(\Delta Y_i - \Delta X_i' \beta - \Delta h_{k(n)}(Z_i) \right)^2 \right. \\ & \times 1\{Y_{i1} \geq \Delta X_i' \beta + \Delta h_{k(n)}(Z_i), Y_{i2} \geq -X_i' \beta - \Delta h_{k(n)}(Z_i)\} + Y_{i1}^2 \\ & \times 1\{Y_{i1} \geq \Delta X_i' \beta + \Delta h_{k(n)}(Z_i), Y_{i2} < -X_i' \beta - \Delta h_{k(n)}(Z_i)\} + Y_{i2}^2 \\ & \left. \times 1\{Y_{i1} < \Delta X_i' \beta + \Delta h_{k(n)}(Z_i), Y_{i2} \geq -X_i' \beta - \Delta h_{k(n)}(Z_i)\} \right] \end{aligned}$$

where $\Delta X_i \equiv X_{i1} - X_{i2}$, $\Delta h(Z_i) \equiv h(Z_{i1}) - h(Z_{i2})$. $\theta_{k(n)}$ enters $R_n(\theta_{k(n)})$ nonsmoothly due to the indicator functions. For notation simplification, we define the following function:

$$\phi(v_1, v_2, \gamma) \equiv \begin{cases} v_1^2, & \text{for } \gamma \leq -v_2, \\ (v_1 - v_2 - \gamma)^2, & \text{for } -v_2 < \gamma < v_1, \\ v_2^2, & \text{for } v_1 \leq \gamma \end{cases}$$

Hence

$$R_n(\theta_{k(n)}) = -\frac{1}{n} \sum_{i=1}^n \phi(Y_{i1}, Y_{i2}, \Delta X_i' \beta + \Delta h_{k(n)}(Z_i))$$

For notation simplicity, we consider one unknown function. The generalization to multiple unknown functions, however, is straightforward. In the appendix A, we propose a two-stage estimation method for the model with multiple unknown functions. The asymptotic properties of the multiple functions model are similar to the single unknown function model and can be easily achieved through the same technique used to draw the results of single unknown function model.

3.3 IDENTIFICATION

Hereafter, we focus on $R_n(\theta)$ to derive identification, consistency and asymptotic normality. We here abuse notation to have $\|\theta\|_s = \|\beta\| + \|h\|_s$, where $\|\cdot\|$ is the Euclidean norm. We will get specific definition for khks below. Recall

$$R_n(\theta_{k(n)}) = -\frac{1}{n} \sum_{i=1}^n \phi(Y_{i1}, Y_{i2}, \Delta X_i' \beta + \Delta h_{k(n)}(Z_i))$$

Let

$$\phi^0(v_1, v_2, \gamma, \gamma_0) = 1\{v_1 > 0, v_2 > 0\}[\phi(v_1, v_2, \gamma) - \phi(v_1, v_2, \gamma_0)]$$

and

$$\begin{aligned} R_n^0(\theta_{k(n)}) &\equiv R_n(\theta_{k(n)}) - R_n(\theta_0) \\ &= -\frac{1}{n} \sum_{i=1}^n \phi^0(Y_{i1}, Y_{i2}, \Delta X_i' \beta + \Delta h_{k(n)}(Z_i), \Delta X_i' \beta_0 + \Delta h_0(Z_i)) \end{aligned}$$

Assumption 3. (i) The parameter space \mathcal{B} is compact, and the true value of the parameter is an interior point of Θ , $\theta_0 \in \text{int } \Theta$; (ii) $h_0(0) = 0$, $h_{k(n)}(0) = 0$ for any $h_{k(n)} \in \mathcal{H}_{k(n)}$ and $h(0) = 0$ for any $h \in \mathcal{H}$.

Remark 1. Our assumption 3 is general. For power series and spline, we can achieve the 0 value by suppressing the intercept term. We can also impose other normalization condition instead, i.e., $\int_{\mathcal{Z}} h_0(z) dz = 0$, $\int_{\mathcal{Z}} h_{k(n)}(z) dz = 0$ and $\int_{\mathcal{Z}} h(z) dz = 0$.

Assumption 4. There is no proper linear subspace of $\mathbb{R}^{d+k(n)}$ containing the random variable $1\{P(Y_{i1} > 0, Y_{i2} > 0 | X_{i1}, X_{i2}, Z_{i1}, Z_{i2})\}(\Delta X'_i, \Delta P_{k(n)}(Z_i))'$ with probability 1.

Lemma 1. If the density of $v_1 - v_2 - \gamma_0$ conditional on $v_1 + v_2$ is strictly unimodal and symmetric around 0, then with assumption 4, $\mathbb{E}(\varphi^0(v_1, v_2, \gamma, \gamma_0))$ achieves its unique minimum at $\gamma = \gamma_0$.

Proof. The proof is the same as that in Lemma A.2 of Honoré (1992).

Proposition 1. If the density of $Y_{i1} - Y_{i2} - \Delta X'_i \beta_0 - \Delta h_0(Z_i)$ conditional on $Y_{i1} + Y_{i2}$ is strictly unimodal and symmetric around $X'_i \beta_0 + \Delta h_0(Z_i) = 0$, with assumption 3 and 4, $\mathbb{E}[R_n^0(\theta_{k(n)})]$ is uniquely minimized at $\theta_{k(n)} = \theta_0$.

Proof. From the lemma 1, we have that $\mathbb{E}[\phi^0(Y_{i1}, Y_{i2}, \Delta X'_i \beta + \Delta h_{k(n)}(Z_i), \Delta X'_i \beta_0 + \Delta h_0(Z_i))]$ is uniquely minimized at a $\theta = (\beta, h_{k(n)})$ such that

$$\Delta X'_i \beta + \Delta h_{k(n)}(Z_i) = \Delta X'_i \beta_0 + \Delta h_0(Z_i) \quad (2)$$

Let

$$h_j^* = \arg \min_{h_{k(n)} \in \mathcal{H}_{k(n)}} \mathbb{E} \left\{ \left(\Delta X'_{i[j]} - \Delta h_{k(n)}(\cdot) \right)^2 \right\}$$

where $X_{i[j]}$ is the j -th element of vector X_i . Denote $h^* = (h_1^*, \dots, h_d^*)'$.

Therefore we have

$$(\Delta X_i - \Delta h^*)' \beta + \Delta h_{k(n)}(Z_i) + \Delta h^{*'} \beta = (\Delta X_i - \Delta h^*)' \beta_0 + \Delta h_0(Z_i) + \Delta h^{*'} \beta_0$$

Left multiplying $(\Delta X_i - \Delta h^*)$ and taking expectation, and with the construction of h^* , we have

$$\mathbb{E}[(\Delta X_i - \Delta h^*)(\Delta X_i - \Delta h^*)'] \beta = \mathbb{E}[(\Delta X_i - \Delta h^*)(\Delta X_i - \Delta h^*)'] \beta_0$$

With full rank of $\mathbb{E}[(\Delta X_i - \Delta h^*)(\Delta X_i - \Delta h^*)']$, we have that $\beta = \beta_0$. Therefore from equation (1), we have

$$\Delta h_{k(n)}(\cdot) = \Delta h_0(\cdot)$$

With assumption 3 and imposing $Z_{i1} = 0$, we have

$$h_{k(n)}(0) - h_{k(n)}(Z_{i2}) = h_0(0) - h_0(Z_{i2})$$

Since Z_{i2} exhausts the support of z , we have that $h_{k(n)} = h_0$. Thus $\mathbb{E}[R_n^0(\theta_{k(n)})]$ is uniquely minimized at $\theta_{k(n)} = \theta_0$.

3.4 CONSISTENCY

Assumption 5. $\mathbb{E}[\|X_t\|^2]$, $\mathbb{E}[\alpha \|\Delta X\|]$, $\mathbb{E}[\|P_{k(n)}(Z_t)\|^2]$, $\mathbb{E}[\alpha \|\Delta P_{k(n)}(Z_t)\|]$, $\mathbb{E}[\epsilon_j \|\Delta X\|]$ and $\mathbb{E}[\epsilon_j \|\Delta P_{k(n)}(Z_t)\|]$ are finite with $j = 1, 2$ and $t = 1, 2$.

Assumption 6. There is a constant $\mu > 0$ such that for any $\theta \in \Theta$, there is $\pi_{k(n)}\theta \in \Theta_{k(n)}$ satisfying $\|\pi_{k(n)}\theta - \theta\|_s = O(k(n)^{-\mu})$ where $k(n)^{-\mu} = o\left(n^{\frac{1}{4}}\right)$. Without loss of generality, we take $\mu = 1$.

Remark 2. We can take $\pi_{k(n)}\theta$ as $\theta_{k(n)}$ while $\pi\theta$ is a general notation.

Remark 3. Assumption 6 is about the approximation error of sieve method. With general series, this assumption holds.

Assumption 7.

$$\ln \left[N \left(\varepsilon^{\frac{1}{K}}, \mathcal{H}_{k(n)}, \|\cdot\|_s \right) \right] \leq \text{const.} \times k(n) \times \ln \frac{k(n)}{\varepsilon}$$

Remark 4. Assumption 7 is the general assumption for the entropy of sieve space, i.e., with $h(\cdot)$ from Donsker class, this assumption holds.

Assumption 8. The conditional density function of X_i and Z_i given Y_i is continuous.

Theorem 1. With assumptions 1-8, the minimizer of $R_n(\theta_{k(n)})$ over $\Theta_{k(n)}$, $\hat{\theta}_{k(n)}$, is a consistent estimator for θ_0 .

Proof. We check the three conditions of Shen and Wong (1994) and apply the Theorem 1 therein to establish consistency. We have

$$\begin{aligned} & \mathbb{E} \left[\phi \left(Y_{i1}, Y_{i2}, \Delta X_i' \beta + \Delta h_{k(n)}(Z_i) \right) - \phi \left(Y_{i1}, Y_{i2}, \Delta X_i' \beta_0 + \Delta h_0(Z_i) \right) \middle| Y_i \right] \\ &= Y_{i1}^2 \{ \Pr(\Delta X_i' \beta + \Delta h_{k(n)}(Z_i) < -Y_{i2} | Y_i) \\ &\quad - \Pr(\Delta X_i' \beta_0 + \Delta h_0(Z_i) < -Y_{i2} | Y_i) \} \\ &\quad + Y_{i2}^2 \{ \Pr(\Delta X_i' \beta + \Delta h_{k(n)}(Z_i) \geq Y_{i1} | Y_i) \\ &\quad - \Pr(\Delta X_i' \beta_0 + \Delta h_0(Z_i) \geq Y_{i1} | Y_i) \} + [Y_{i1} - Y_{i2} - \Delta X_i' \beta + \Delta h_{k(n)}(Z_i)] \\ &\quad \times \Pr(-Y_{i2} < \Delta X_i' \beta + \Delta h_{k(n)}(Z_i) < Y_{i1} | Y_i) \\ &\quad - [Y_{i1} - Y_{i2} - \Delta X_i' \beta_0 + \Delta h_0(Z_i)] \\ &\quad \times \Pr(-Y_{i2} < \Delta X_i' \beta_0 + \Delta h_0(Z_i) < Y_{i1} | Y_i) \end{aligned}$$

Since cumulative density function is continuous in θ , we have that $|F(\theta_{k(n)}; \cdot) - F(\theta_0; \cdot)| \leq A_1 \|\theta_{k(n)} - \theta_0\|_s$. Furthermore we have that $G(\theta_{k(n)}; \cdot) \equiv [Y_{i1} - Y_{i2} - \Delta X'_i \beta - \Delta h_{k(n)}(Z_i)]^2$ is continuous in $\theta_{k(n)}$ and therefore we have similar result that $|G(\theta_{k(n)}; \cdot) - G(\theta_0; \cdot)| \leq A_2 \|\theta_{k(n)} - \theta_0\|_s$. Constants $A_1, A_2 > 0$ and $\|\theta_{k(n)} - \theta_0\|_s \equiv \|\beta - \beta_0\| + \|\mathbf{h}_{k(n)} - \mathbf{h}_0\|_s$. After adding and subtracting $[Y_{i1} - Y_{i2} - \Delta X'_i \beta_0 + \Delta h_{k(n)}(Z_i)] \times \Pr(-Y_{i2} < \Delta X'_i \beta_0 + \Delta h_0(Z_i) < Y_{i1} | Y_i)$ in above equation, and with finiteness of Y_{it} , we have

$$\begin{aligned} & \mathbb{E} \left[\phi(Y_{i1}, Y_{i2}, \Delta X'_i \beta + \Delta h_{k(n)}(Z_i)) - \phi(Y_{i1}, Y_{i2}, \Delta X'_i \beta_0 + \Delta h_0(Z_i)) \middle| Y_i \right] \\ & \leq \{A_1(Y_{i1}^2 + Y_{i2}^2) + A_2[Y_{i1} - Y_{i2} - \Delta X'_i \beta_0 + \Delta h_0(Z_i)]^2 + 1\} \times \\ & \quad \|\theta_{k(n)} - \theta_0\|_s \end{aligned}$$

With finite support of Y_i , we have that

$$\begin{aligned} & \mathbb{E} \left[\phi(Y_{i1}, Y_{i2}, \Delta X'_i \beta + \Delta h_{k(n)}(Z_i)) - \phi(Y_{i1}, Y_{i2}, \Delta X'_i \beta_0 + \Delta h_0(Z_i)) \middle| Y_i \right] \leq A \times \\ & \quad \|\theta_{k(n)} - \theta_0\|_s, \quad A < \infty \end{aligned}$$

Therefore, Condition C1 in Shen and Wong (1994) holds with $\alpha = \frac{1}{2}$.

Similarly, we can easily show that

$$\begin{aligned} & \text{Var} \left[\phi(Y_{i1}, Y_{i2}, \Delta X'_i \beta + \Delta h_{k(n)}(Z_i)) - \phi(Y_{i1}, Y_{i2}, \Delta X'_i \beta_0 + \Delta h_0(Z_i)) \middle| Y_i \right] \leq B \times \\ & \quad \|\theta_{k(n)} - \theta_0\|_s, \quad B < \infty \end{aligned}$$

Thus Condition C2 in Shen and Wong (1994) holds with $\beta = 1$ (here β and above α are following the definition of those in Shen and Wong (1994)). Condition C3 holds with assumption 7. Thus we have

$$\|\hat{\theta}_{k(n)} - \theta_0\|_s = O_p(\max\{n^{-\tau}, \|\pi_{k(n)}\theta_0 - \theta_0\|_s, K(\pi_{k(n)}\theta_0, \theta_0)\}) = o_p(1)$$

where $K(\pi_{k(n)}\theta_0, \theta_0) \equiv \mathbb{E}[\phi(\cdot, \theta_0) - \phi(\cdot, \pi_{k(n)}\theta_0)]$ and $\tau = \frac{1}{2} - \frac{\ln \ln(n)}{\ln(n)}$

[Assumption 7 indicates $r_0 = \frac{1}{4}$ and $r = 0^+$ in Shen and Wong (1994)].

Remark 5. *With series selection, we can make $K(\pi_{k(n)}\theta_0, \theta_0)$ small enough. $\|\pi_{k(n)}\theta_0 - \theta_0\|_s$ is the approximation error and is small enough. In general, if the rate of $K(\pi_{k(n)}\theta_0, \theta_0)$ and $\|\pi_{k(n)}\theta_0 - \theta_0\|_s$ are functions of some known nuisance parameters, we can draw the best rate of $\|\hat{\theta}_{k(n)} - \theta_0\|_s$.*

3.5 ASYMPTOTIC NORMALITY

Define

$$\begin{aligned} & \phi'_{\theta_0}(\cdot)[\theta_{k(n)} - \theta_0] \\ & \equiv \begin{cases} 0, & \Delta D_0 \leq -Y_{i2}, \\ -2(\Delta Y_i - \Delta X_i'\beta_0 - \Delta h_0(Z_i))[\Delta X_i'(\beta - \beta_0) + \Delta h_{k(n)} - \Delta h_0], & -Y_{i2} < \Delta D_0 < Y_{i1}, \\ 0, & \Delta D_0 \geq Y_{i1}. \end{cases} \end{aligned}$$

where $\Delta D_0 = \Delta X_i'\beta_0 + \Delta h_0(Z_i)$, and

$$\begin{aligned} & r(\cdot)[\theta_{k(n)} - \theta_0] \\ & \equiv \phi(Y_{i1}, Y_{i2}, \Delta X_i'\beta + \Delta h_{k(n)}(Z_i)) - \phi(Y_{i1}, Y_{i2}, \Delta X_i'\beta_0 + \Delta h_0(Z_i)) \\ & \quad - \phi'_{\theta_0}(\cdot)[\theta_{k(n)} - \theta_0] \end{aligned}$$

Now we proceed to check the conditions in Shen (1997) to establish the asymptotic normality of $\hat{\theta}_{k(n)}$. We verify them one by one below.

Condition A:

$$\sup_{\theta_{k(n)} \in \Theta_{k(n)} : \|\theta_{k(n)} - \theta_0\|_s \leq \delta_n} \frac{1}{\sqrt{n}} v\{r(\cdot)[\theta_{k(n)} - \theta_0] - r(\cdot)[\pi_{k(n)}\theta^* - \theta_0]\} = O_p(\zeta_n^2)$$

where $\theta^* = (1 - \zeta_n)\theta + \zeta_n(\theta_0 + u^*)$, $\zeta_n = o\left(n^{-\frac{1}{2}}\right)$, $u^* = \pm v^*$ and v^* is the Riesz representer which will be defined below. $v(g) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n [g(\cdot) - \mathbb{E}_0 g(\cdot)]$. Because $r(\cdot)[\cdot]$ is nonzero only in the middle support, we here focus on these nonzero values $r(\cdot)[\cdot]$.

$$\begin{aligned} r(\cdot)[\theta_{k(n)} - \theta_0] - r(\cdot)[\pi_{k(n)}\theta^* - \theta_0] \\ = r(\cdot)[\theta_{k(n)} - \theta_0] - r(\cdot)[\theta^* - \theta_0] + r(\cdot)[\theta^* - \theta_0] \\ - r(\cdot)[\pi_{k(n)}\theta^* - \theta_0] \end{aligned}$$

and

$$\begin{aligned} r(\cdot)[\theta_{k(n)} - \theta_0] - r(\cdot)[\theta^* - \theta_0] \\ = [\Delta X'_i(\beta - \beta_0) + \Delta h_{k(n)}(Z_i) - \Delta h_0(Z_i)]^2 \\ - [\Delta X'_i(\beta^* - \beta_0) + \Delta h^*(Z_i) - \Delta h_0(Z_i)]^2 \\ = [\Delta X'_i(\beta - \beta_0) + \Delta h_{k(n)}(Z_i) - \Delta h_0(Z_i) + \Delta X'_i(\beta^* - \beta_0) + \Delta h^*(Z_i) \\ - \Delta h_0(Z_i)] \times [\Delta'_i(\beta - \beta^*) + \Delta h_{k(n)}(Z_i) - \Delta h^*(Z_i)] \\ \approx O_p(\|\theta_{k(n)} - \theta_0\|_s) \times O_p(\|\theta_{k(n)} - \theta^*\|_s) = O_p(\zeta_n \times \|\theta_{k(n)} - \theta_0\|_s) \end{aligned}$$

Similarly, when $\|\theta_{k(n)} - \theta_0\|_s \leq \delta_n$, we have

$$r(\cdot)[\theta^* - \theta_0] - r(\cdot)[\pi_{k(n)}\theta^* - \theta_0] = O_p(\zeta_n^2).$$

Thus when we take $\delta_n = O(\zeta_n)$

$$\begin{aligned}
& \sup_{\theta_{k(n)} \in \Theta_{k(n)} : \|\theta_{k(n)} - \theta_0\|_s \leq \delta_n} \frac{1}{\sqrt{n}} v\{r(\cdot)[\theta_{k(n)} - \theta_0] - r(\cdot)[\pi_{k(n)}\theta^* - \theta_0]\} \\
&= \frac{1}{n} \sum_{i=1}^n \{r(\cdot)[\theta_{k(n)} - \theta_0] - r(\cdot)[\pi_{k(n)}\theta^* - \theta_0]\} = \frac{1}{n} \times n \times O_p(\zeta_n^2) \\
&= O_p(\zeta_n^2)
\end{aligned}$$

Therefore Condition A in Shen (1997) holds in our case.

Condition B:

$$\begin{aligned}
& \sup_{\theta_{k(n)} \in \Theta_{k(n)} : 0 < \|\theta_{k(n)} - \theta_0\|_s \leq \delta_n} [K(\pi_{k(n)}\theta^*, \theta_0) - K(\theta_{k(n)}, \theta_0)] \\
& - \frac{1}{2} [\|\theta^* - \theta_0\|_s^2 - \|\theta_{k(n)} - \theta_0\|_s^2] = O(\zeta_n^2)
\end{aligned}$$

Since function $K(\cdot, \cdot)$ is an expectation, it is second differentiable in θ . Define

$$\|\theta_{k(n)} - \theta_0\|_s \equiv \left[(\theta_{k(n)} - \theta_0)' \frac{d^2 K(\theta_{k(n)}, \theta_0)}{d\theta d\theta'} (\theta_{k(n)} - \theta_0) \right]^{\frac{1}{2}}$$

Thus

$$\begin{aligned}
K(\theta^*, \theta_0) &= K(\theta_0, \theta_0) + \frac{dK(\theta_0, \theta_0)}{d\theta} (\theta^* - \theta_0) + \frac{1}{2} (\theta^* - \theta_0)' \frac{d^2 K(\theta^*, \theta_0)}{d\theta d\theta'} (\theta^* - \theta_0) \\
&+ \text{small term}
\end{aligned}$$

We have $K(\theta_0, \theta_0) = 0$ and $\frac{dK(\theta_0, \theta_0)}{d\theta} = 0$. Hence we have

$$\begin{aligned}
K(\pi_{k(n)}\theta^*, \theta_0) &= \frac{1}{2} \|\pi_{k(n)}\theta^* - \theta_0\|_s^2 + o_p(\|\pi_{k(n)}\theta^* - \theta_0\|_s) \\
&= \frac{1}{2} (\|\pi_{k(n)}\theta^* - \theta^*\|_s^2 + 2 \|\pi_{k(n)}\theta^* - \theta^*\|_s \|\theta^* - \theta_0\|_s + \\
&\quad \|\theta^* - \theta_0\|_s^2) + o_p(\|\pi_{k(n)}\theta^* - \theta_0\|_s) \\
&= O(\delta_n^{-2} \zeta_n^4) + O(\delta_n^{-1} \zeta_n^2) \times \delta_n + \frac{1}{2} \|\theta^* - \theta_0\|_s^2 + o(\zeta_n^2) = \frac{1}{2} \times \\
&\quad \|\theta^* - \theta_0\|_s^2 + O(\zeta_n^2)
\end{aligned}$$

Thus we have Condition B satisfied.

Condition C:

$$\sup_{\theta_{k(n)} \in \Theta_{k(n)}: 0 < \|\theta_{k(n)} - \theta_0\|_s \leq \delta_n} \|\theta^* - \pi_{k(n)}\theta^*\|_s = O(\delta_n^{-1} \zeta_n^2)$$

Condition C holds with conventional sieve estimators, i.e. Power series, Fourier Series, Spline series, etc.

With Assumption 7 and the Lemma 4 in Shen and Wong (1994), we have that Condition D in Shen (1997) holds in our case. Define $f(\theta) \equiv \beta$, we have

Theorem 2

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, V)$$

where $V = \text{Var}_0(\phi'_{\theta_0}(\cdot)[v^*])$ and v^* satisfying $\langle v^*, \theta_{k(n)} - \theta_0 \rangle = f'_{\theta_0}(\theta_{k(n)} - \theta_0) = \beta - \beta_0$.

3.6 CONCLUSION

In this paper, we generalize the linear panel data truncated model with fixed effect to allow for a nonparametric component. With primitive assumptions, we identify the

parametric parameter and the nonparametric unknown function. We propose a sieve estimation method for our nonstandard truncated regression model and establish the consistencies of both parametric parameter and nonparametric unknown function. Furthermore, we achieve \sqrt{n} convergent rate for the parametric parameter. Economic theories seldom suggest parametric relationships between economic variables. Therefore, semiparametric models are more desirable when we have less evidence on specific parametric relationship. With our approach, we can avoid misspecification problem and still have the power to give explanation for economic questions with the parametric component estimation. While we establish the \sqrt{n} asymptotic normality for the parametric parameter, the asymptotic normality of the nonparametric component lacks in this paper. For this part, we refer to Horowitz and Lee (2005) for further extension.

CHAPTER IV

A SEMIPARAMETRIC APPROACH TO ESTIMATE THE DYNAMIC
EFFECTIVENESS OF MARKETING MIX IN MARKET RESPONSE MODELS
IN PIONEER-FOLLOWER CONTEXT, WITH VENKATESH SHANKAR

4.1. INTRODUCTION

Market response models describe how outputs, sales or market shares, respond to the marketing activities and how competitors interact. The responsiveness of marketing mix variables change over the product life cycle [Parsons (1975)] and also depend on contextual variables. Therefore the corresponding marketing strategy and tactics should vary over different stages and contexts [Parsons (1975), Gatignon and Hanssens (1987), Parker (1992)]. With appropriately estimated market response models, firms could optimally allocate the resource among different marketing mix instruments. The illustration of marketing mix interactions serve as a basis for marketing strategies for marketing mix resource allocation [Gatignon and Hanssens (1987)]. How to model the interaction mechanism is crucial to investigate the relationship between market performance and marketing efforts [Gatignon and Hanssens (1987)]. The interaction mechanism serves like a black box hence the structure assumptions become important for market response models. Multiplicative model is widely used in market response analyses due to its convenience to position marketing mix variables and contextual variables separately. Though multiplicative model provides systematic knowledge about the determinants of marketing responsiveness, it suffers from several drawbacks. First, it

usually has a linear marketing parameter function which may cause misspecification problem. Second, there are usually collinearity problems in the data in the multiplicative form. In this paper, we provide a flexible semiparametric model which avoids the misspecification and near multicollinearity problems.

Though managers like a response rule for every period and environment, the responsiveness of output, i.e. sales or market shares, with respect to inputs, i.e., advertising, promotion, detailing, etc., usually are varying and depend on the context in which the response happens. There are two type variables: input variables, i.e. marketing mix, and stimuli variables, i.e. contextual variables [Bowman and Gatignon (2009)]. How to model these two type variables appropriately is the key to the market response model. The basic model frequently used in the literature is the multiplicative model. Multiplicative model provides an intuitive way to deal with the two type variables by putting the contextual variables in the coefficients. We deploy terminologies marketing response function and marketing parameter function in our analyses which is similar as those used in Gatignon and Hanssens (1987). The marketing parameter function describes the way in which marketing responsiveness is generated. Therefore, the contextual variables enter in the marketing parameter function. The market response function illustrates how market performance is achieved and hence the marketing mix variables are major components. Extant literature either takes the marketing parameter function as constant or as linear combination of contextual variables [Parsons (1975), Gatignon and Hanssens (1989), Bowman and Gatignon (1996), etc.] An exception is Parker (1992) who allows quadratic time term in the coefficients. Though multiplicative

model has good performance in market response estimations, it suffers from the misspecification and near multicollinearity problems. We introduce a flexible semiparametric model to avoid misspecification and near multicollinearity problems. We allow the market parameter functions to be unknown functions of contextual variables and use nonparametric kernel approach to estimate these unknown functions. The specific definitions of response function and parameter function will be described in details in section 4.

In pharmaceutical industry, the competition is not price-based, therefore we investigate the responsiveness of advertising and detailing, as described in Gatignon, Weitz and Bansal (1990). Competitive environment, i.e. industry concentration, firm size and market familiarity all affect the responsiveness of marketing mix variables and the entry strategy. The responsiveness of marketing mix variables are affected by firm's capabilities and market conditions [Gatignon, Weitz and Bansal (1990)]. The former leads to specific product quality level and the latter impacts on the relative marketing effort, the ratio of marketing expenditure. Therefore firm capabilities, market context, and relative product quality have a moderating effect on the relationship between relative marketing effort and the performance of a brand. In section 3, we discuss the determinants of responsiveness in details while due to data restriction; we will only use the order of entry and the time in the market as moderators. The number of competitors has high correlation with competitive marketing expenditure and hence we drop it.

The competitive reaction to the marketing mix may offset the effectiveness of these variables and therefore change the responsiveness shape of marketing mix instruments.

The order of entry does not only contribute to market share/sales through main effect, but also indirectly serve as moderator of marketing mix responsiveness [Bowman and Gatignon (1996)]. Whether the followers can catch up the pioneer through marketing effort is important. How does the order of entry affect the responsiveness of marketing mix instruments leads to different marketing strategy to allocate the resource? Is there asymmetry among responsiveness for the pioneer and the followers? Due to the near multicollinearity problem, the main effect of the order of entry dominates in the parametric model. Furthermore, in the more flexible semiparametric model, the main effect of order of entry is still significant, not minimal as documented in Bowman and Gatignon (1996). To overcome the disadvantage of entering later, the followers have to spend substantially more money.

In our semiparametric estimation, the late entrants have higher responsiveness for advertising than detailing, which suggests the late entrants allocate more resource in advertising in the early stage to avoid order-of-entry effect, though the responsiveness of detailing is usually higher than the corresponding responsiveness of advertising. While in the late stage, the high responsiveness of advertising for the late entrant diminishes and the late entrants are better to spend more money in detailing. The responsiveness of competitive expenditure is not statistically significant. Table 9 provides literature review and comparison of market responsive model and related topics.

Table 9 Literature on Market Response Models

Paper	Marketing Mix Variables	Contextual Variables	Model
Parsons (1975)	Advertising	Time	Multiplicative model with exponential marketing parameter function
Gatignon and Hanssens (1987)	Advertising; Number of Recruiters	Average Propensity; Local Advertising Support	Multiplicative Model with linear marketing parameter function
Parker (1992)	Price	Time	Diffusion Model with quadratic elasticity function
Bowman and Gatignon (1996)	Price; Promotion; Advertising; Product Quality; Competitive Advertising	Order of Entry; Number of Competitors	Multiplicative Model with linear marketing parameter function
Shankar and Bayus (2003)	Price; Advertising	Network Size; Product Quality;	Multiplicative Model with linear marketing parameter function
Valratsas, Feinberg, Bass and Kalyanram (2004)	Advertising; Distribution; Price; Order of Entry; Time in Market	Regime Threshold	Multiplicative Model with Regime Switch
Van Heerde, Mela, And Manchanda (2004)	Price; Promotion;	Innovation Brand: Pioneer Dummy and Early Follower Dummy	Multiplicative Model with linear auto-regressive marketing parameter function
Narayanan, Manchanda, and Chintagunta (2005)	Marketing Communication/Detailing	Time	Structural Model of Demand and Random Coefficient Models with a Bayesian learning process
Danaher, Bonfrer, and Dhar (2008)	Price ; Advertising; Promotion	Competitive Clutter	Multiplicative model with exponential marketing parameter function
Osinga, Leeftang, and Wieringa (2010)	Detailing; Competitive Marketing Expenditure	Detailing; Competitive Marketing Expenditure	Structural Multiplicative Model
Lin and Shankar (2013)	Advertising; Detailing; Competitive Marketing Expenditure	Order of Entry; Time in Market	Smooth Coefficient Multiplicative Model

Table 10 gives a brief comparison of parametric and semiparametric methods in several dimensions. Our out-of-sample results confirm the well-established advantages

and drawbacks of semiparametric methods. With rich data, the semiparametric methods usually outperform the parametric methods while the performance reverses with small sample size [Horowitz (2009), Li and Racine (2007)].

Table 10 Comparison of Parametric and Semiparametric Models

Comparison of Parametric and Semiparametric Models		
Dimension	Parametric Model	Semiparametric Model
Small Sample Size	Better	Worse
Rich Data	Worse	Better
Dynamic Responsiveness	Fail	Good Managerial Implications
Model Fitting	Worse	Better

Our paper proceeds as follows. In section 2, we briefly discuss the data used in the analysis. Section 3 investigates the determinants of responsiveness. The models are developed in section 4 and the corresponding estimation methods are proposed. Section 5 illustrates the results from different model specifications and we draw the main conclusions. The managerial implications from conventional parametric model and the smooth varying coefficient model are discussed in section 6. Section 7 provides several limitations of our model and future research directions. The semiparametric smooth varying coefficient model is introduced in the appendix C.

4.2. DATA

Data from two pharmaceutical categories, Fluoxetine and Statin, are used in our analyses. The data for Fluoxetine is from Jan 1988 to Dec 1995 and the time range for Statin is from Oct 1987 to Jan 1996. In both categories, we choose the first four brands, one pioneer and three early followers. We will use all four firms for our market response analyses of different approaches while in the out of sample prediction analyses, we will use part of the data due to time windows selection. We have monthly advertising, detailing and total prescriptions for all top four brands. The descriptive statistics of the variables appear in Table 11. Firms invest more in detailing than in advertising.

Table 11 Summary Statistics of Key Variables

Fluoxetine			Statin		
Firm	Var.	Mean (Std)	Firm	Var.	Mean (Std)
PROZAC	AD	98.15(85.75)	MEVACOR	AD	202.40(231.68)
	DT	2062.43(618.12)		DT	1573.74(453.18)
	TP	888.48(364.52)		TP	659.46(269.73)
ZOLOFT	AD	541.39(394.93)	PRAVACHOL	AD	481.31(864.62)
	DT	2719.89(754.91)		DT	1878.88(732.87)
	TP	678.50(314.37)		TP	313.27(134.06)
PAXIL	AD	469.8(502.27)	ZOCOR	AD	287.66(434.53)
	DT	2906.51(1250.92)		DT	1822.02(366.47)
	TP	469.06(206.00)		TP	304.13(194.93)
LUVOX	AD	258.22(116.71)	LESCOL	AD	761.00(522.96)
	DT	1533.11(309.81)		DT	2085.70(908.53)
	TP	41.56(9.95)		TP	227.95(107.19)

In our analysis, we pool the four firms' data and take advertising, detailing, competitive expenditure as the marketing mix. Because the number of competitors is highly correlated with logarithm competitive marketing expenditure, we drop it from the contextual variables. Therefore we have the order of entry and the time in the market as moderators. In this paper, we use “contextual variables” and “moderator” interchangeably. Table 2 provides literature review on market response model. In those papers, both marketing mix variables and contextual variables differ.

4.3. DETERMINANTS OF RESPONSIVENESS

4.3.1. Order of Entry

How does the order of entry enter the response model? Literature suggests two main functions: main effect or moderator. Bowman and Gatignon (1996) investigate the role of the order of entry. They suggest that the order of entry is more like a moderator. The benefits of early entry are that the pioneer can choose the most profitable segment of the market and also the pioneer could claim monopolistic profits. The costs of early entry include facing uncertainty and the resistance from consumers to adopt new product, i.e., Oracle named its first product ORACLE 2.0, actually the first edition, is one attempt to reduce the resistance of buyers to purchase the first edition (Larry Ellison).

4.3.2. Number of Competitors

Number of competitors captures the competitive contextual effect. Because competitors usually initiate corresponding marketing strategies and this reduces the effectiveness of marketing mix expenditure [Steenkamp, Nijs, Hanssens and Dekimpe (2005)]. The competitive reactions from competitors offset the efforts made and the magnitude of competitive reactions is highly correlated with the number of competitors. Therefore the number of competitors has contextual effect on the market output.

4.3.3. Time in the Market

The diffusion literature suggests in different stages of the product, the effectiveness of the marketing mix instruments varies. Generally, in the early stage of new product, the exposure is important to attract early adopters and therefore advertising and detailing play key roles at the early stage of the new product. During the matured stage, usually promotion takes the role from advertising and detailing since the follower-type consumers are more price sensitive. Therefore the time of a product in the market is important for the analysis of responsiveness and the time in the market serves as one contextual variable in market response model. Bass diffusion model and its extensions provide details for the effect of stages on market performance.

4.3.4. Network Size

Customer network is argued to be an asset of firms and the network size affects the responsiveness of marketing mix expenditure [Shankar and Bayus (2003)]. In the

technological categories, the compatibility is very important. Products of Apple always suffer from the incompatibility issue though they have large sales and market shares. The abundance of app store somewhat affects the competitions among main smart phone manufacturers and the effectiveness of marketing mix instruments.

4.3.5. Innovation

The innovation plays a similar role as network size. There are several platform competitions behind the terminal product competitions. For instance, there are three main smart phone operation systems, iOS, Android and Windows'. Competitions among smart phones usually are impacted by which platforms they deployed. Therefore the market responsiveness is affected by innovation [van Heerde, Mela and Manchanda (2004)].

4.4. MODEL DEVELOPMENT AND ESTIMATION

4.4.1. Parametric Multiplicative Model Specification

We first put our analysis in the well-received multiplicative model and illustrate two main drawbacks it suffers.

$$\begin{aligned} \ln TP_{it} = & \alpha(OE_{it}, NC_{it}, TM_{it}) + \ln AD_{it-1} \beta_1(OE_{it}, NC_{it}, TM_{it}) \\ & + \ln DT_{it-1} \beta_2(OE_{it}, NC_{it}, TM_{it}) + \ln CME_{it-1} \beta_3(OE_{it}, NC_{it}, TM_{it}) \\ & + \epsilon_{it} \end{aligned} \tag{1}$$

where

$$\begin{aligned}
\alpha(OE_{it}, NC_{it}, TM_{it}) &= \alpha_0 + \alpha_1 OE_{it} + \alpha_2 NC_{it} + \alpha_3 TM_{it} \\
\beta_1(OE_{it}, NC_{it}, TM_{it}) &= \beta_{10} + \beta_{11} OE_{it} + \beta_{12} NC_{it} + \beta_{13} TM_{it} \\
\beta_2(OE_{it}, NC_{it}, TM_{it}) &= \beta_{20} + \beta_{21} OE_{it} + \beta_{22} NC_{it} + \beta_{23} TM_{it} \\
\beta_3(OE_{it}, NC_{it}, TM_{it}) &= \beta_{30} + \beta_{31} OE_{it} + \beta_{32} NC_{it} + \beta_{33} TM_{it}
\end{aligned}$$

The first drawback of parametric multiplicative model is the near multicollinearity problem. The following table documents the correlation among the marketing mix variables, the contextual variables, and their interaction terms. We denote that two variables are highly correlated if their correlation is larger than 75%. Those highly correlated pairs are marked yellow in Table 12 and Table 13. From Table 12, we know that only $\ln AD$, $\ln DT$, $\ln CME$, TM and OE are free from near multicollinearity. Therefore we only keep these variables and model (1) reduces to

$$\begin{aligned}
\ln TP_{it} &= \alpha_0 + \alpha_1 OE_{it} + \alpha_3 TM_{it} + \beta_{10} \ln AD_{it-1} + \beta_{20} \ln DT_{it-1} + \beta_{30} \ln CME_{it-1} \\
&+ \epsilon_{it}
\end{aligned} \tag{2}$$

In model (2), the contextual variables, OE and TM enter the market response model as main effects. Therefore, the late entrants can spend money in AD and DT to overcome the disadvantage of not entering as the pioneer while take advantage of the benefits of entering later, less uncertainty and less consumer resistance towards adopting a new innovation. However, the contextual variables usually not only have the main effect, they also serve as moderators on the responsiveness of marketing mix variables.

Therefore, due to near multicollinearity issue, the linear multiplicative may not be suitable for capturing such moderator role of contextual variables. For the STATIN category, the bear multicollinearity problem is slightly weaker than that of FLUOXETINE category, see Table 13. Though the near multicollinearity problem is relative weak, we can also exclude a lot of regressors from model (1) and have

$$\begin{aligned}\ln TP_{it} = & \alpha_0 + \alpha_1 OE_{it} + \alpha_3 TM_{it} + \beta_{10} \ln AD_{it-1} + \beta_{20} \ln DT_{it-1} + \beta_{30} \ln CME_{it-1} \\ & + \beta_4 \ln AD_{it-1} \times \ln CME_{it} + \epsilon_{it}\end{aligned}\quad (2')$$

We only have six regressors in the regression.

Table 12 Correlation Matrix for Model (1): Fluoxetine

Correlation Matrix of Marketing Mix Variables, Contextual Variables and their interaction terms for FLUOXETINE																		
	$\ln AD$	$\ln DT$	$\ln CME$	TM	OE	NC	$\ln AD \times \ln DT$	$\ln AD \times \ln CME$	$\ln DT \times \ln CME$	$\ln AD \times TM$	$\ln AD \times OE$	$\ln AD \times NC$	$\ln DT \times TM$	$\ln DT \times OE$	$\ln DT \times NC$	$\ln CME \times TM$	$\ln CME \times OE$	$\ln CME \times NC$
$\ln AD$	1.00																	
$\ln DT$	0.49	1.00																
$\ln CME$	0.30	0.50	1.00															
TM	-0.35	0.22	0.23	1.00														
OE	0.46	0.17	0.51	-0.55	1.00													
NC	0.30	0.40	0.89	0.26	0.58	1.00												
$\ln AD \times \ln DT$	0.99	0.58	0.37	-0.31	0.46	0.34	1.00											
$\ln AD \times \ln CME$	0.66	0.55	0.89	-0.02	0.62	0.81	0.71	1.00										
$\ln DT \times \ln CME$	0.33	0.54	1.00	0.23	0.50	0.88	0.40	0.91	1.00									
$\ln AD \times TM$	-0.06	0.28	0.29	0.91	-0.44	0.36	-0.03	0.16	0.29	1.00								
$\ln AD \times OE$	0.73	0.32	0.50	-0.54	0.93	0.54	0.73	0.73	0.50	-0.37	1.00							
$\ln AD \times NC$	0.57	0.44	0.82	0.06	0.66	0.93	0.60	0.90	0.82	0.26	0.72	1.00						
$\ln DT \times TM$	-0.33	0.24	0.25	1.00	-0.54	0.27	-0.29	0.01	0.25	0.91	-0.53	0.08	1.00					
$\ln DT \times OE$	0.51	0.27	0.54	-0.53	0.99	0.60	0.53	0.66	0.54	-0.41	0.95	0.69	-0.51	1.00				
$\ln DT \times NC$	0.32	0.43	0.89	0.26	0.57	1.00	0.37	0.83	0.89	0.37	0.54	0.94	0.28	0.59	1.00			
$\ln CME \times TM$	-0.20	0.29	0.58	0.89	-0.29	0.56	-0.15	0.33	0.58	0.84	-0.29	0.36	0.90	-0.26	0.57	1.00		
$\ln CME \times OE$	0.46	0.31	0.75	-0.33	0.95	0.77	0.48	0.79	0.74	-0.23	0.89	0.81	-0.32	0.96	0.76	-0.01	1.00	
$\ln CME \times NC$	0.29	0.39	0.88	0.26	0.58	1.00	0.33	0.80	0.87	0.36	0.54	0.93	0.27	0.59	1.00	0.56	0.76	1.00

Table 13 Correlation Matrix of Model (1): Statin

Correlation Matrix of Marketing Mix Variables, Contextual Variables and their interaction terms for STATIN																		
	$\ln AD$	$\ln DT$	$\ln CME$	TM	OE	NC	$\ln AD \times \ln DT$	$\ln AD \times \ln CME$	$\ln DT \times \ln CME$	$\ln AD \times TM$	$\ln AD \times OE$	$\ln AD \times NC$	$\ln DT \times TM$	$\ln DT \times OE$	$\ln DT \times NC$	$\ln CME \times TM$	$\ln CME \times OE$	$\ln CME \times NC$
$\ln AD$	1.00																	
$\ln DT$	-0.06	1.00																
$\ln CME$	-0.09	0.29	1.00															
TM	-0.36	-0.12	0.21	1.00														
OE	0.08	0.25	0.50	-0.52	1.00													
NC	-0.15	0.22	0.92	0.27	0.55	1.00												
$\ln AD \times \ln DT$	1.00	0.02	-0.07	-0.38	0.11	-0.13	1.00											
$\ln AD \times \ln CME$	0.74	0.13	0.57	-0.19	0.40	0.47	0.76	1.00										
$\ln DT \times \ln CME$	-0.10	0.38	1.00	0.19	0.51	0.91	-0.06	0.56	1.00									
$\ln AD \times TM$	0.51	-0.35	0.00	0.46	-0.35	0.01	0.47	0.39	-0.04	1.00								
$\ln AD \times OE$	0.74	0.17	0.26	-0.51	0.64	0.27	0.75	0.78	0.27	0.04	1.00							
$\ln AD \times NC$	0.70	0.08	0.55	-0.14	0.47	0.56	0.71	0.95	0.54	0.40	0.82	1.00						
$\ln DT \times TM$	-0.38	-0.06	0.22	1.00	-0.52	0.28	-0.40	-0.20	0.20	0.42	-0.51	-0.16	1.00					
$\ln DT \times OE$	0.09	0.33	0.50	-0.53	1.00	0.55	0.12	0.40	0.52	-0.37	0.65	0.47	-0.51	1.00				
$\ln DT \times NC$	-0.15	0.31	0.92	0.25	0.57	1.00	-0.13	0.47	0.92	-0.03	0.28	0.55	0.26	0.57	1.00			
$\ln CME \times TM$	-0.37	-0.01	0.56	0.90	-0.28	0.59	-0.39	0.04	0.54	0.34	-0.36	0.07	0.90	-0.28	0.57	1.00		
$\ln CME \times OE$	0.03	0.30	0.73	-0.35	0.96	0.74	0.06	0.50	0.73	-0.28	0.59	0.55	-0.34	0.96	0.75	-0.04	1.00	
$\ln CME \times NC$	-0.16	0.23	0.92	0.27	0.54	1.00	-0.14	0.46	0.91	0.00	0.26	0.54	0.28	0.54	0.99	0.59	0.74	1.00

4.4.2. Smooth Varying Coefficient Model Specification

Here we introduce a flexible model which avoids the near multicollinearity naturally by its model structure, taking $\alpha(\cdot), \beta_1(\cdot), \beta_2(\cdot)$ and $\beta_3(\cdot)$ as unknown functions.

$$\begin{aligned} \ln TP_{it} = & \alpha(OE_{it}, TM_{it}) + \ln AD_{it-1} \beta_1(OE_{it}, TM_{it}) + \ln DT_{it} \beta_2(OE_{it}, TM_{it}) \\ & + \ln CME_{it} \beta_3(OE_{it}, TM_{it}) + \varepsilon_{it} \end{aligned} \quad (3)$$

where $\alpha(\cdot), \beta_1(\cdot), \beta_2(\cdot)$ and $\beta_3(\cdot)$ are unknown smooth functions to be estimated nonparametrically. The smooth varying coefficient model overcomes another disadvantage of the parametric model, the endogeneity issue comes from misspecification. Rewrite the model (1) as

$$Y_{it} = X'_{it} \gamma(Z_{it}) + \varepsilon_{it} \text{ where } \gamma(Z_{it}) = Z'_{it} \delta \quad (4)$$

and

$$Y_{it} = X'_{it} \gamma_0(Z_{it}) + \varepsilon_{it} \text{ where } \gamma_0(Z_{it}) \neq Z'_{it} \delta \cdots \text{True model} \quad (5)$$

To simplify the illustration, we assume that ε_{it} is exogenous. Thus

$$\varepsilon_{it} = X'_{it} [\gamma_0(Z_{it}) - \gamma(Z_{it})] + \varepsilon_{it}$$

which is correlated with X_{it} and the endogeneity issue arises. Though smooth varying coefficient model also has an endogenous problem due to the correlation between X_{it} and ε_{it} , while it avoids such an endogeneity issue caused by misspecification by reducing misspecification probability with flexible model setting up. Therefore our semiparametric smooth varying coefficient model mitigates the two problems in the parametric multiplicative model. Beside these two main model advantages, we will

document other merits of our model flexibility in terms of estimation results, i.e., R^2 , model fitting, and out-of-sample analysis.

4.4.3. Model Estimation

We estimate the model (2) and (3) with standard parametric and semiparametric approaches. The detailed estimation procedure for the semiparametric smooth varying coefficients model is in the Appendix. In the semiparametric specification, we use ad hoc bandwidth selection. We draw our conclusions from the estimations in the next section. In models above, we exclude the price because of the characteristics of the pharmaceutical industry.

4.5. RESULTS

4.5.1. Estimation Results

We get our estimation results for the parametric and semiparametric multiplicative model specifications in Table 6 and Table 7. We focus on the parametric and semiparametric specifications. Since the coefficients estimations for the semiparametric specification are varying and we will use figures to illustrate the responsiveness below. For the Fluoxetine category, there is strong order of entry effect, that there is strong pioneer advantage. All coefficients are significant in the parametric model in Fluoxetine category. However, the results from the parametric model for the Statin category are different. First, there is pioneer disadvantage which combined with the pioneer advantage for Fluoxetine illustrates the benefit and drawback of entering earlier provided

that the models are correct. Furthermore, the coefficients for detailing and competitive marketing expenditure are insignificant. The more flexible the model specification is, the larger the R^2 is. This indicates that more flexible model specifications increase the model fitting performance which can be found in Figure 4 and 5.

Table 14 Estimation Results: Fluoxetine

Estimation Results: Fluoxetine Category		
Variable	Parametric Model	Semiparametric Model
Intercept	2.133***	
$\ln AD_{it}$	0.146***	
$\ln DT_{it}$	0.443***	
$\ln CME_{it}$	-0.042**	
TM_{it}	0.023***	
OE_{it}	-0.277**	
R^2	0.7136	0.9412893

*10% significant; ** 5% significant; *** 10% significant

Table 15 Estimation Results: Statin

Estimation Results: Statin Category		
Variable	Parametric Model	Semiparametric Model
Intercept	4.646 ***	
$\ln AD_{it}$	0.139**	
$\ln DT_{it}$	-0.054	
$\ln CME_{it}$	-0.018	
TM_{it}	0.033***	
OE_{it}	0.165***	
$\ln AD_{it-1} \times \ln CME_{it}$	-0.013*	
R^2	0.6854	0.8989212

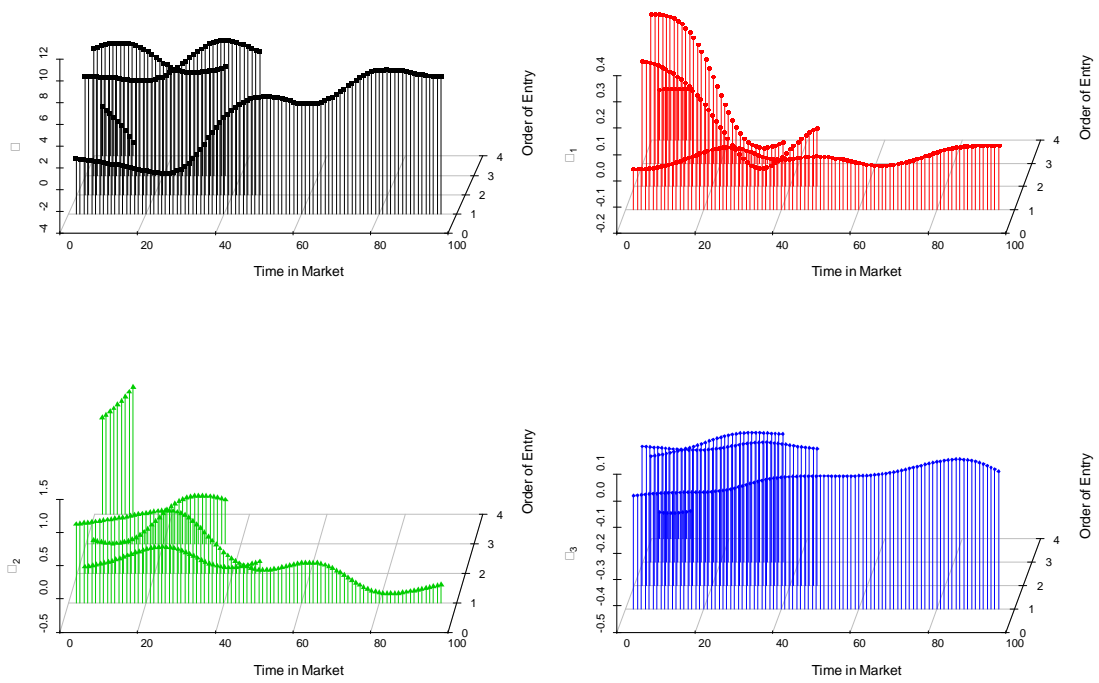


Figure 4 Smooth Varying Responsiveness of Fluoxetine Category

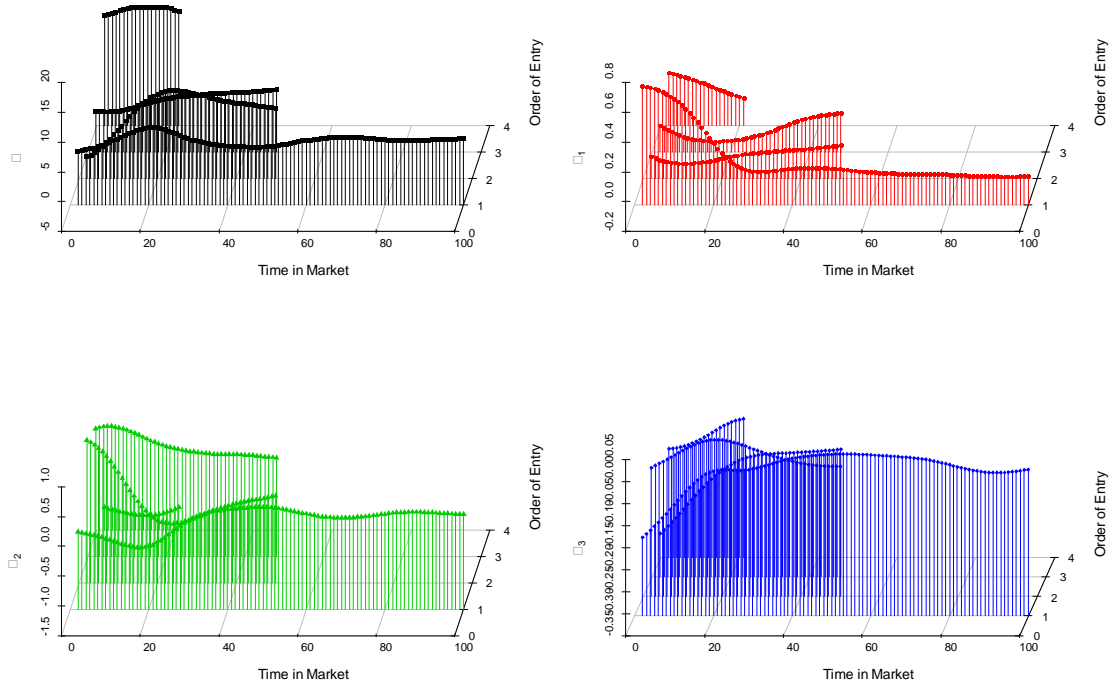


Figure 5 Smooth Varying Responsiveness of Statin Category

The smooth varying responsiveness is illuminated in Figure 4 and 5. We see obvious varying effectiveness of advertising and detailing. We will explore the details of these varying coefficients in section 6 about the managerial implications. The model fitting comparison can be found in Figure 6 and Figure 7. We separate the fitting for top four brands. In terms of R^2 and model fitting, our results coincide with the intuition that the more flexible is the model specification, the better performance is the model.

The model fittings show that the semiparametric fitted data are more closed to the real data than the parametric multiplicative fitted data. This evidence supports the result

that the performance in terms of R^2 or model fitting decreases from semiparametric specified model to the parametric multiplicative model.

4.5.2. Out-of-Sample Analysis

In this subsection, we use different time range data and the brands in the corresponding range to investigate in which scenario the semiparametric model is better and in which condition it is worse than the parametric model. We use three-year to seven-year time windows to estimate the in-sample model and with another half-year data for out-of-sample predictions. The squares of prediction errors are shown in Table 16.

In Fluoxetine Category, PROZAC enters as pioneer, ZOLOFT enters 50 months later, PAXIL enters 61 months later and LUVOX enters 85 month later. Therefore for the first four years, there is only one pioneer brand: PROZAC. When choosing a five-year frame, there are two brands. There are three brands for more than six years.

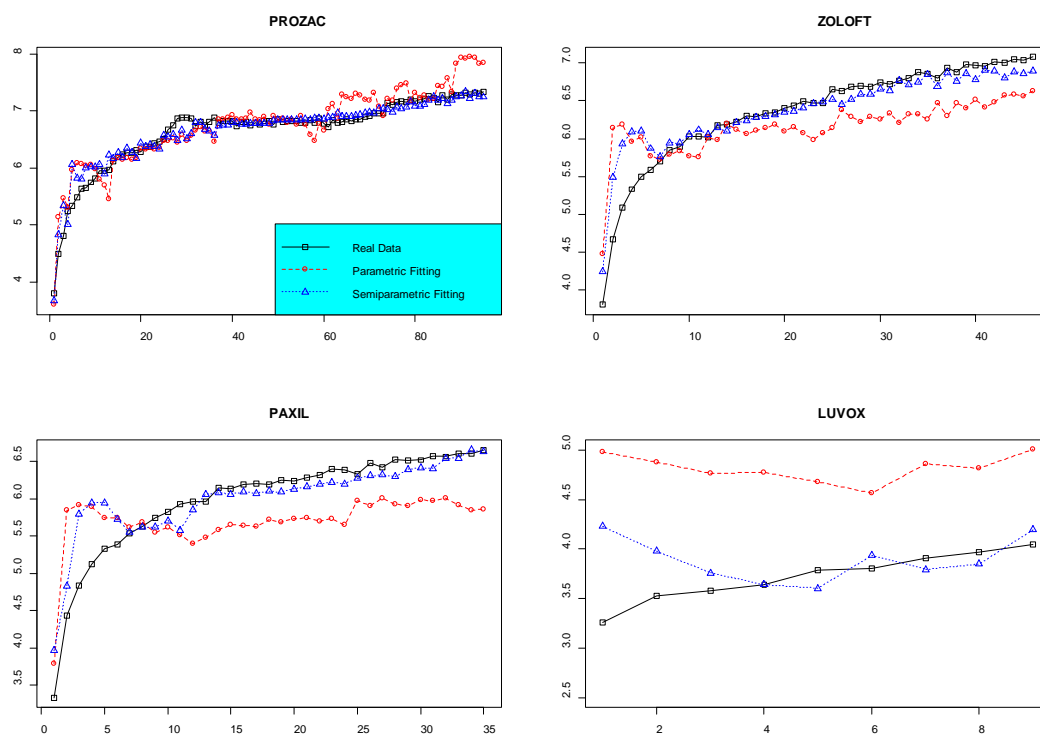


Figure 6 Model Fitting Comparison of Fluoxetine Category

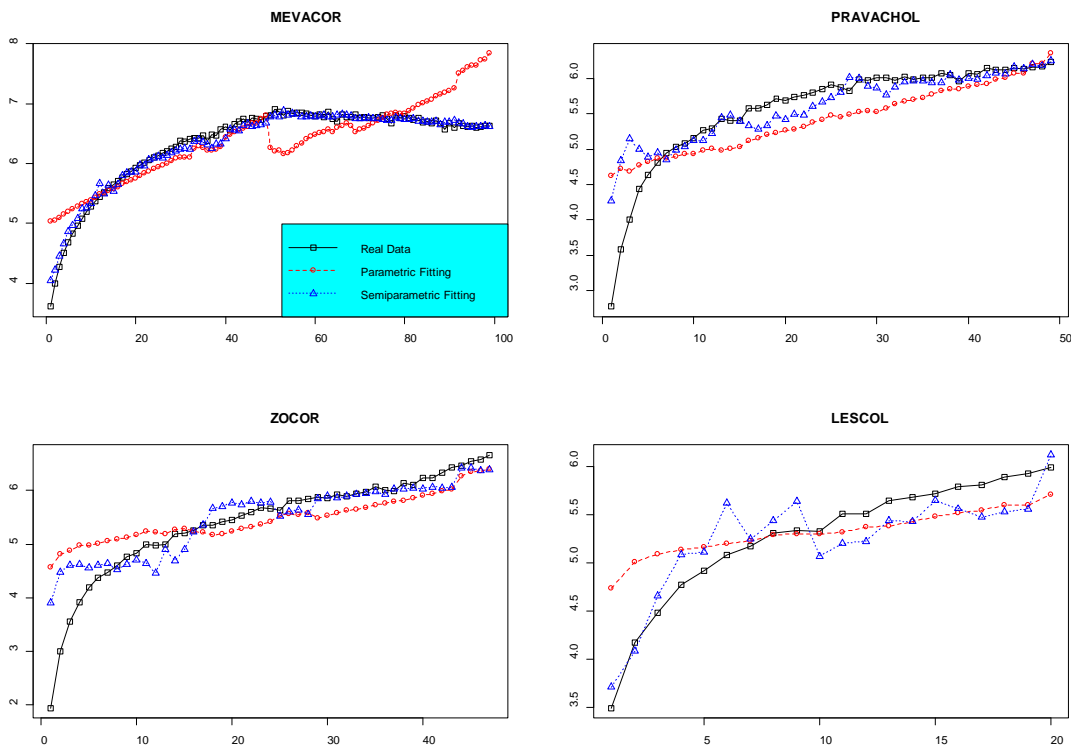


Figure 7 Model Fitting Comparison of Statin Category

Table 16 Square of Prediction Errors with Different Time Windows

		Square of Prediction Errors: Fluoxetine				
Time Window		3-Year	4-Year	5-Year	6-Year	7-Year
Parametric	One-Brand	1.02	1.13			
Semiparametric	One-Brand	0.01	0.01			
Parametric	Two-Brand			1.93		
Semiparametric	Two-Brand			2.00		
Parametric	Three-Brand				3.23	2.74
Semiparametric	Three-Brand				1.84	1.08
		Square of Prediction Errors: Statin				
Time Window		3-Year	4-Year	5-Year	6-Year	7-Year
Parametric	One-Brand	0.94	1.31			
Semiparametric	One-Brand	0.16	0.05			
Parametric	Two-Brand					
Semiparametric	Two-Brand					
Parametric	Three-Brand			7.20	5.22	4.48
Semiparametric	Three-Brand			7.39	1.35	0.27

The results listed above are the square of prediction errors for different time-frames. From Table 16, we observe that the out-of-sample performance improves when the data sample size increases for the semiparametric approach.

We also conclude that when there is not enough data, the parametric model outperforms the semiparametric specification with the prediction figures 8 and 9 for Fluoxetine category and Figure 10 and 11 for Statin category in appendix B for each brand in all seven time frames. Figure 8 provides out-of-sample predictions with three-year, four-year and five year time frames. For the first two time frames, the semiparametric model predicts better than the parametric model while the results reverse with for the first year of ZOLOFT. This result is confirmed with the first year performances of the parametric and semiparametric model in Figure 9. The results for the two followers are similar as that of the pioneer. Therefore the semiparametric model has advantages in richer data scenario while the parametric model should be used with small sample sized data. In Statin Category, MEVACOR enters as pioneer, PRAVACHOL enters 51 months later, PAXIL enters 53 months later and LUVOX enters 80 month later. The results are similar as those in Fluoxetine.

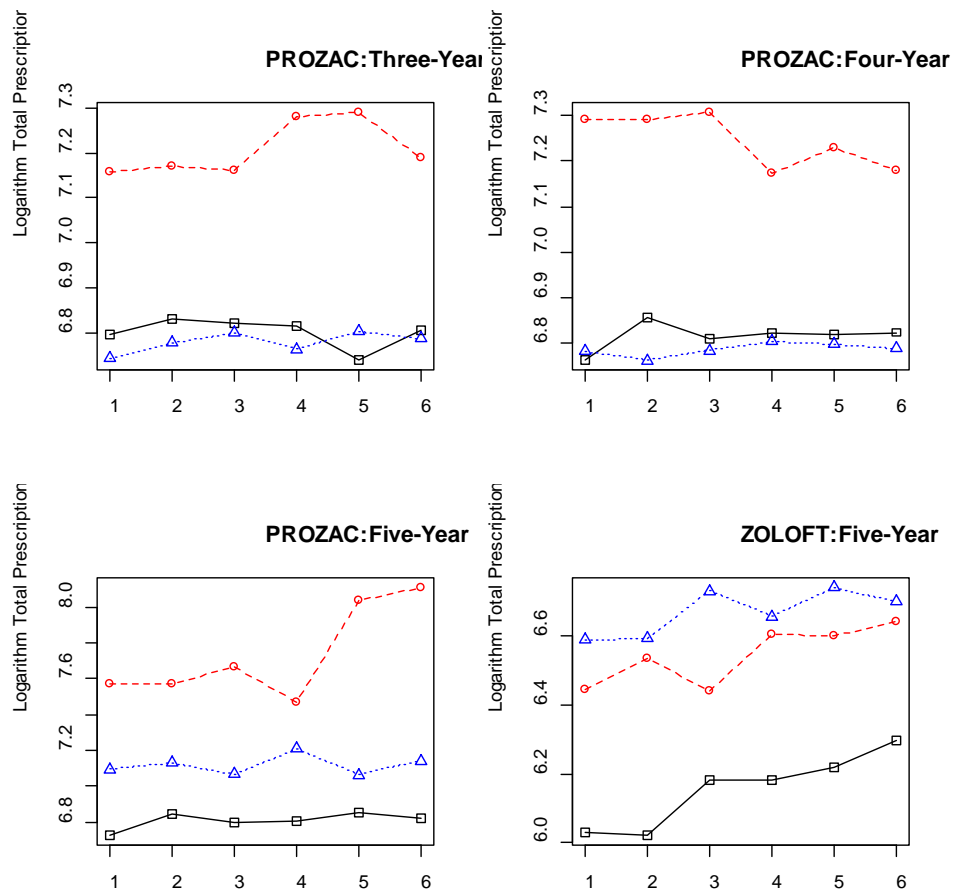


Figure 8 Out-of-Sample Predictions with Three to Five Years “In Sample” Time Windows: Fluoxetine

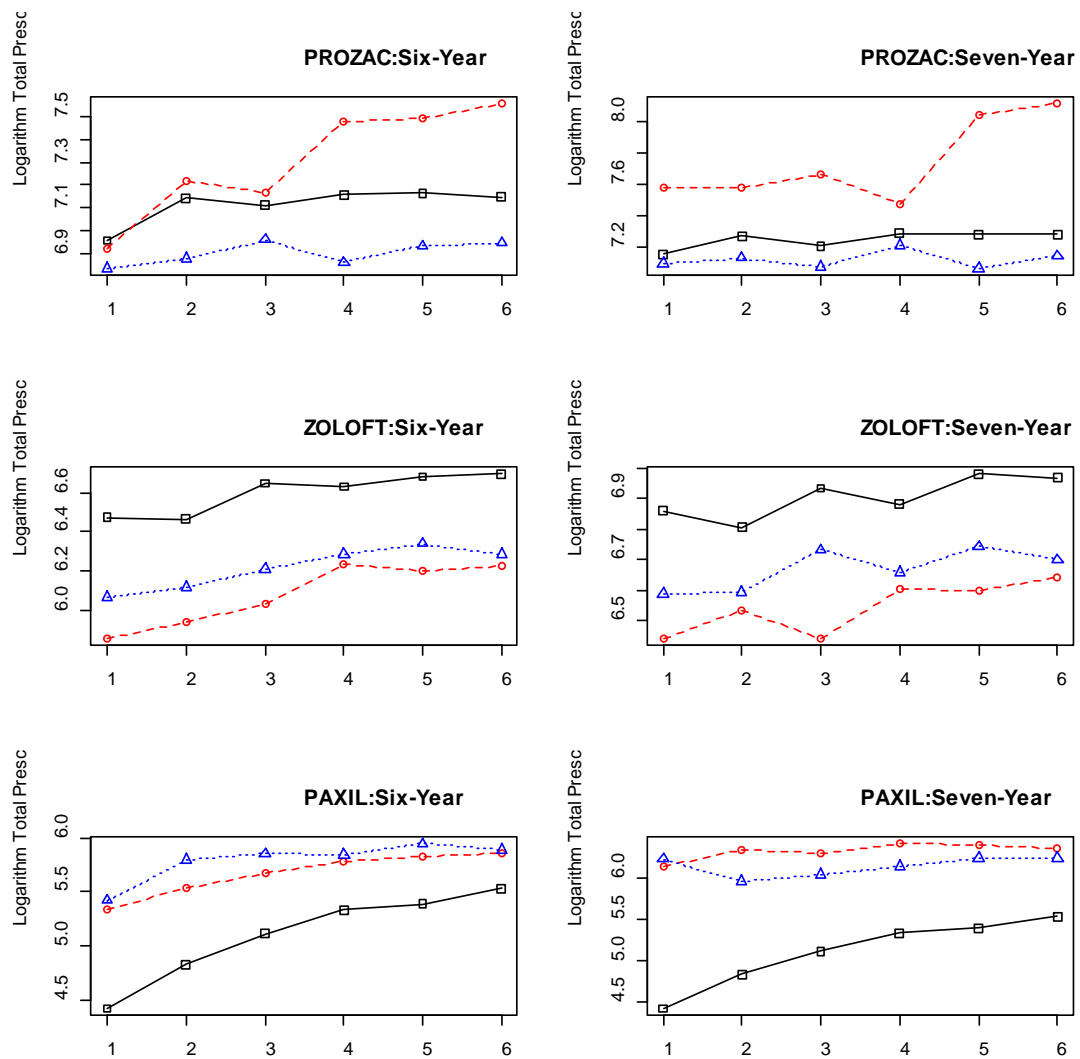


Figure 9 Out-of-Sample Predictions with Six to Seven Years “In Sample” Time Windows: Fluoxetine

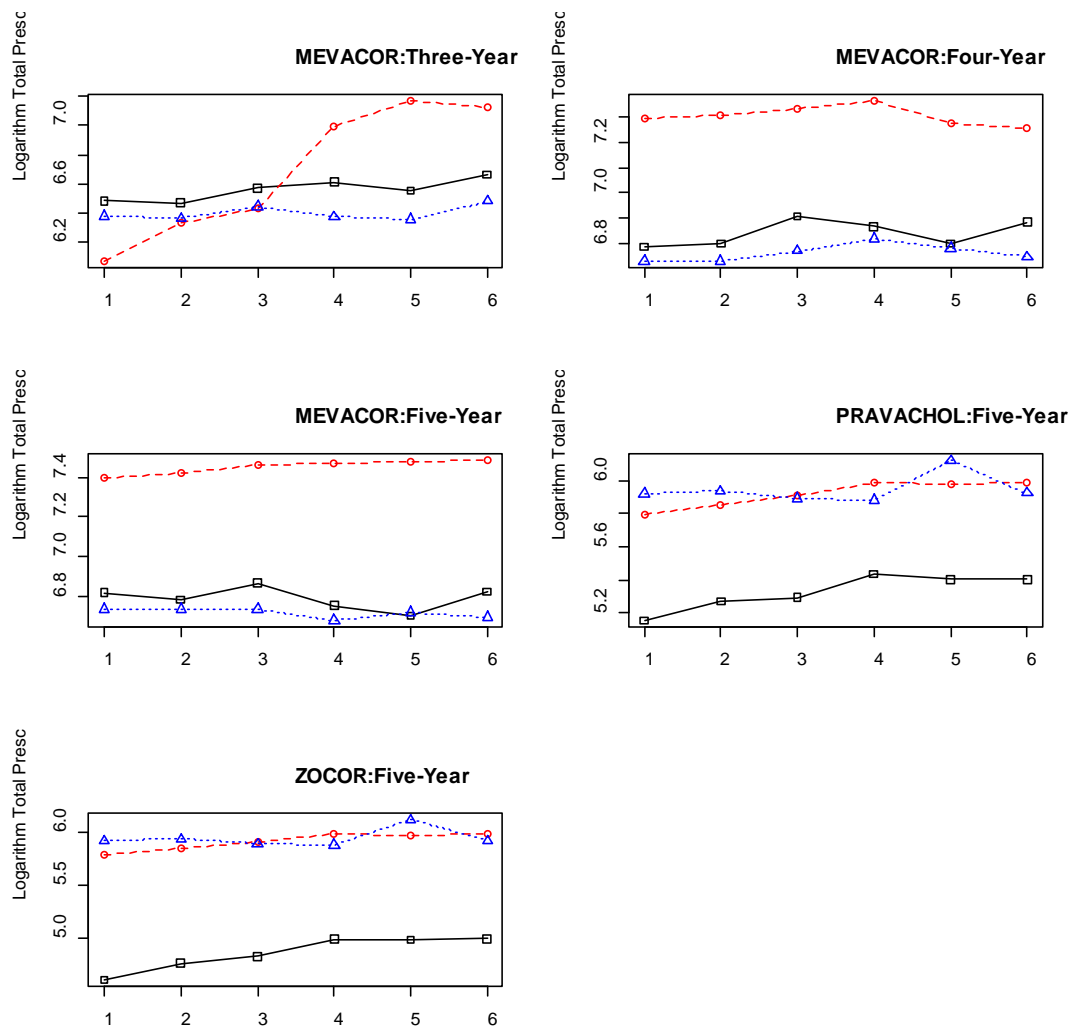


Figure 10 Out-of-Sample Predictions with Three to Five Years “In Sample” Time Windows: Statin

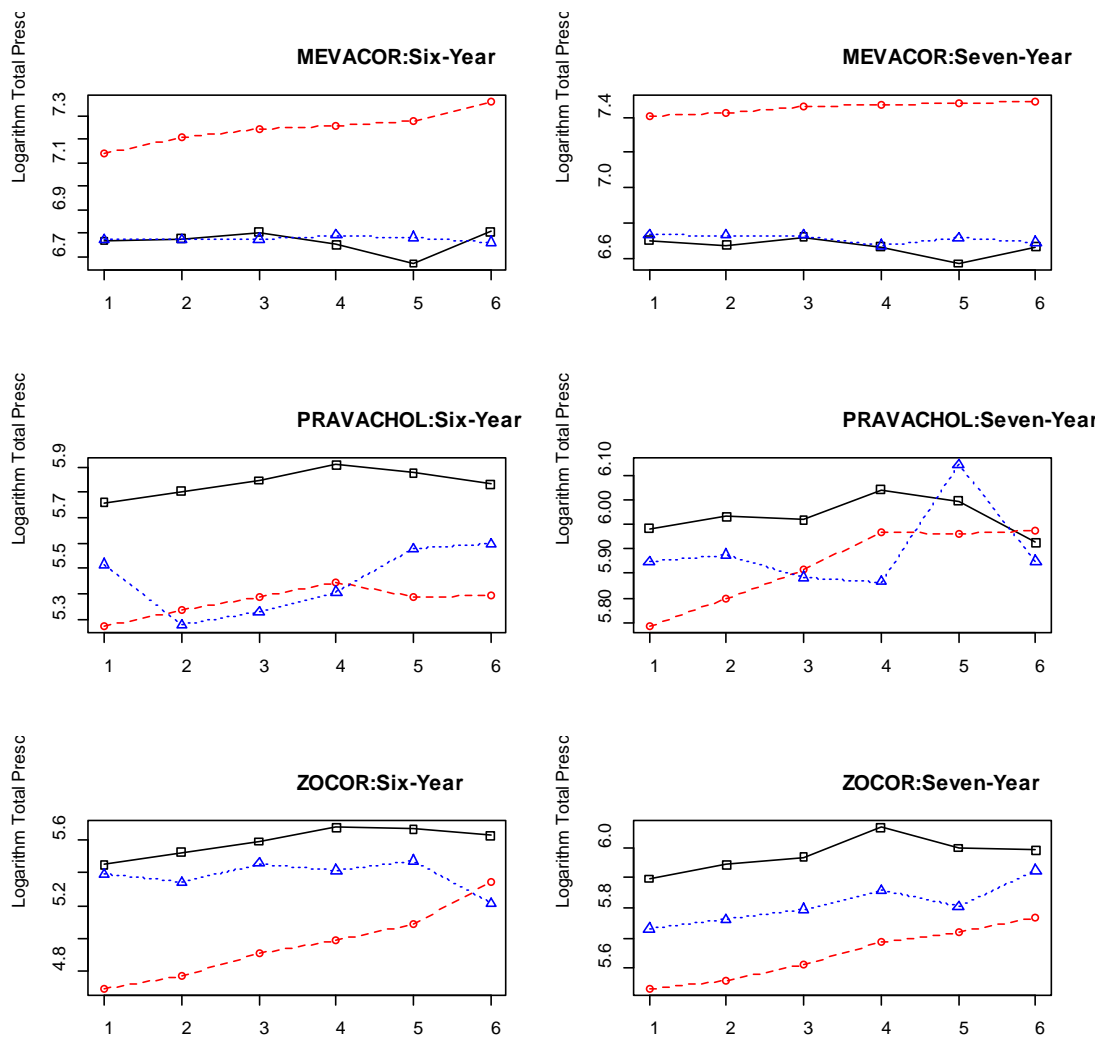


Figure 11 Out-of-Sample Predictions with Six to Seven Years “In Sample” Time Windows: Statin

4.6. MANAGERIAL IMPLICATIONS

4.6.1. Implications Based on the Parametric Multiplicative Model Estimation

From Table 14, we learn that advertising and detailing contribute to the sales of the product and the effect of detailing is much larger than that of advertising. This suggests firms to allocate more resource to detailing. This is intuitive for the pharmaceutical industry since the prescriptions are made by physicians who are specialists and detailing provides better targeting. There is a slight but significant up trend for the total prescriptions. The longer the brand is in the market, the higher is the sale. This may due to the growth of the Fluoxetine category. Our results also document the general sense that the advantages of early entry dominate the disadvantages while the order of entry in our model conveys the main effect. Therefore late entrants could offset the later entering by investing more in advertising and detailing. Also it is not uncommon that the competitive marketing expenditure has negative impacts on the sale. This documents the competitive effect even though we do not have the number of competitors.

From Table 15, we learn that in Statin category, the firms should invest more in advertising rather than in detailing. This may be misleading because it is well known that detailing has higher effectiveness in pharmaceutical industry.

4.6.2. Implications Based on Smooth Varying Coefficient Model Estimation

The implications derived from the smooth varying coefficient model estimation are much dynamic and reasonable than the static ones from parametric multiplicative model. We here look at the varying responsiveness brand by brand. From Figure 4, in the early

stage, the pioneer has higher responsiveness for detailing than that of advertising and hence the pioneer should spend more money in detailing. As the other brands enter the market, the responsiveness of detailing for the pioneer vanishes while the positive responsiveness of advertising for the pioneer persists. Therefore, at the competitive stage, the pioneer should allocate more resource to advertising. For the followers, the story is reverse. At their early entry stage, due to existing incumbent, the followers have relative advantages in advertising and these advantages disappear as the time goes on. Therefore the allocation rules for the followers should be in the opposite direction as that of the pioneer.

For the Statin category, the story is different. From Figure 8, we observe that the pioneer, during the early stage of product life cycle, the responsiveness of advertising is stronger while at the late stage the effectiveness of detailing is higher. This provides the managerial implication that MEVACOR is better off if it allocates more resource to advertising at the pioneer stage and invests more in detailing during the competitive stage.

4.7. CONCLUSION

Our analysis is limited to two pharmaceutical categories and the conclusions drawn may be different in other categories or other industries while extensions of our model could be conducted in future. Furthermore, we have the order of entry and the time in the market as contextual variables due to data restrictions. Future research can be done with richer datasets by investigating the impact of network size, innovation and other

contextual variables. In our model, the sample sizes are 96 and 100 for the two pharmaceutical categories which may affect the performances of the smooth varying coefficient model. Because the semiparametric models usually need more data than the parametric model due to its local utilization of the data. The superiority of the smooth varying coefficient model could be documented stronger with richer dataset in the future.

CHAPTER V

CONCLUSIONS

We have belief that all effects are context-based, e.g. peer effects strength depends on the status of interacting individuals, competitive effects among firms depends on the order of entry. This belief amounts to econometrics is the semiparametric varying coefficient models. This dissertation contributes to the literature in theoretical and empirical perspective. We provide theoretical analysis to discrete game model with hierarchy and status and panel data semiparametric truncated regression model with fixed effects. With theoretical results for discrete game model, we illustrate the findings with empirical study of peer effects in college attendance among high school students. Also, we provide dynamic analysis of the effectiveness of advertising and detailing in pharmaceutical industry and find different pattern of effectiveness for pioneer and follower firms in different stages of competition.

REFERENCES

- Aguirregabiria, V., and P. Mira (2007): Sequential Estimation of Dynamic Discrete Games, *Econometrica* 75(1), 1-53.
- Akerlof, G.A. (1997): Social Distance and Social Decisions, *Econometrica* 65(5), 1005-1027.
- Andrews, D. (1991): Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models, *Econometrica* 59, 307-345.
- Angrist, J.D., and K. Lang (2004): Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program, *American Economic Review* 94(5): 1613-1634.
- Badev, A. (2013): Discrete Games in Endogenous Networks: Theory and Policy, Working Paper.
- Bajari, P., H. Hong, J. Krainer, and J. Nekipelov (2010): Estimating Static Models of Strategic Interactions, *Journal of Business and Economic Statistics* 28(4), 469-482.
- Bala, V., and S. Goyal (2000): A Noncooperative Model of Network Formation, *Econometrica* 68(5), 1181-1229.
- Ball, S., C. Eckel, P.J. Grossman, and W. Zame (2001): Status in Markets, *Quarterly Journal of Economics* 116(1), 161-187.
- Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006): Who's Who in Networks. Wanted: The Key Player, *Econometrica* 74(5), 1403-1417.
- Bault, N., G. Coricelli, and A. Rustichini (2008): Interdependent Utilities: How Social Ranking Affects Choice Behavior, *PLoS One* 3(10), 1-10.

- Bayer, P., R. Hjalmarsson, and D. Pozen (2009): Building Criminal Capital behind Bars: Peer Effects in Juvenile Corrections, *Quarterly Journal of Economics* 124(1): 105-147.
- Becker, G.S., K.M. Murphy, and I. Werning (2005): The Equilibrium Distribution of Income and the Market for Status, *Journal of Political Economy* 113(2), 282-310.
- Blume, L.E, W.A. Brock, S.N. Durlauf, and Y.M. Ioannides (2011): Identification of Social Interactions, *Handbook of Social Economics*, 853-964.
- Blume, L. E, W.A. Brock, S.N. Durlauf, and R. Jayaraman (2013): Linear Social Interactions Models, *National Bureau of Economic Research*.
- Bonacich, P. (1987): Power and Centrality: A Family of Measures, *American Journal of Sociology* 92(5), 1170-1182.
- Bowman, D. and H. Gatignon (1996): Order of Entry as a Moderator of the Effect of the Marketing Mix on Market Share, *Marketing Science* 15, 222-242.
- Bowman, D. and H. Gatignon (2009): Market Response and Marketing Mix Models: Trends and Research Opportunities, *Foundations and Trends in Marketing* 4(3), 129-207.
- Brock, W.A., and S.N Durlauf (2001): Discrete Choice with Social Interactions, *Review of Economic Studies* 68(2), 235-260.
- Cai, Z. and Q. Li (2008): Nonparametric Estimation of Varying Coefficient Dynamic Panel Data Models, *Econometric Theory* 24(5), 1321-1342.
- Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009): Peer Effects and Social Networks in Education, *Review of Economic Studies* 76, 1239-1267.

- Carrell, S.E., M. Hoekstra, and J.E. West (2011): Is Poor Fitness Contagious? Evidence from Randomly Assigned Friends, *Journal of Public Economics* 95, 657-663.
- Carrell, S.E., B.I. Sacerdote, and J.E. West (2013): From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation, *Econometrica* 81(3), 855-882.
- Chen, X. (2007): Large Sample Sieve Estimation of Semi-Nonparametric Models, *Handbook of Econometrics* 6B, 5549-5632.
- Christakis, N.A., J.H. Fowler, G.W. Imbens, and K. Kalyanaraman (2010): An Empirical Model for Strategic Network formation, Working Paper.
- Danaher, P.J., A. Bonfrer and S. Dhar (2008): The Effect of Competitive Advertising Interference on Sales for Packaged Goods, *Journal of Marketing Research* 45(2), 211-225.
- Dekimpe, M.G. and D.M. Hanssens (1995): The Persistence of Marketing Effects on Sales, *Journal of Marketing Research* 14(1), 1-21.
- Duflo, E., and E. Saez (2003): The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment, *Quarterly Journal of Economics* 118(3), 815-842.
- Durlauf, S.N., and Y.M. Ioannides (2010): Social Interactions, *Annual Review of Economics* 2, 451-478.
- Ellison, G. (1993): Learning, Local Interaction, and Coordination, *Econometrica* 61(5), 1047-1071.
- Fogli, A., and L. Veldkamp (2013): Germs, Social Networks and Growth, working paper.

- Galeotti, A., S. Goyal, M.O. Jackson, F. Vega-Redondo, and L. Yariv (2010): Network Games, *Review of Economic Studies* 77, 218-244.
- Gatignon, H., E. Anderson and K. Helsen (1989): Competitive Reactions to Market Entry: Explaining Interfirm Differences, *Journal of Marketing Research* 26(1), 44-55.
- Hiroiyuki, K., and K. Shimotsu (2012): Sequential Estimation of Structural Models with a Fixed Point Constraint, *Econometrica* 80(5), 2303-2319.
- Honoré, B. (1992): Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects, *Econometrica* 60, 533-565.
- Honoré, B., and J. Powell (1992): Pairwise Difference Estimators of Censored and Truncated Regression Models, *Journal of Econometrics* 64, 241-278.
- Horowitz, J.L. (2009): Semiparametric and Nonparametric Methods in Econometrics. New York: Springer.
- Horowitz, J.L., and S. Lee (2002): Semiparametric Methods in Applied Econometrics: Do the Models Fit the Data? *Statistical Modelling* 2, 3-22.
- Horowitz, J.L., and S. Lee (2005), Nonparametric Estimation of an Additive Quantile Regression Model, *Journal of the American Statistical Association* 100, 1238-1249.
- Horowitz, J.L., and C. Manski (1995), Identification and Robustness with Contaminated and Corrupted Data, *Econometrica* 63(1), 281-302.
- Horowitz, J.L., and C. Manski (2006), Identification and Estimation of Statistical Functionals Using Incomplete Data, *Journal of Econometrics* 132(2), 445-459.

- Hurwicz, L. (1950): Generalization of the Concept of Identification, Statistical Inference in Dynamic Economic Models. New York: John Wiley & Sons, Inc.
- Jackson, M.O., and A. Wolinsky (1996): A Strategic Model of Social and Economic Networks, *Journal of Economic Theory* 71(1), 44-74.
- Jackson, M.O., and L. Yariv (2007): Diffusion Behavior and Equilibrium Properties in Network Games, *American Economic Review* 97(2), 92-98.
- Katz, L. (1953): A New Status Index Derived from Sociometric Analysis, *Psychometrika* 18, 39- 43.
- Kline, B. (2012): Identification of Complete Information Games, working paper.
- Koopmans, T.C., and O. Reiersøl (1950): The Identification of Structural Characteristics, *Annals of Mathematical Statistics* 21, 165-181.
- Lewbel, A., and O. Linton (2002): Nonparametric Censored and Truncated Regression, *Econometrica* 70(2), 765-779.
- Li, Q. and J. Racine (2007): Nonparametric Econometrics: Theory and Practice. Princeton, NJ: Princeton University Press.
- Luttmer, E.F.P. (2005): Neighbors as Negatives: Relative Earnings and Well-Being, *Quarterly Journal of Economics* 120(3), 963-1002.
- Magee, J.C., and A.D. Galinsky (2008): Social Hierarchy: The Self-Reinforcing Nature of Power and Status, *The Academy of Management Annals* 2(1), 351-398.
- Manski, C. (1995): Identification Problems in the Social Sciences. Cambridge, MA: Harvard University Press.

- Manski, C.F. (2000): Economic Analysis of Social Interactions, *Journal of Economic Perspective* 14(3), 115-136.
- Manski, C. (2003): Partial Identification of Probability Distributions. New York: Springer-Verlag.
- Manski, C. (2005): Partial Identification with Missing Data: Concepts and Findings, *International Journal of Approximate Reasoning* 39, 151-165.
- Manski, C.F. and D.A.Wise (1983): College Choice in America. Cambridge, MA: Harvard University Press.
- Mele, A. (2011): A Structural Model of Segregation in Social Networks, Working Paper.
- Molinari, F. (2008): Partial Identification of Probability Distributions with Misclassified Data, *Journal of Econometrics* 144(1), 81-117.
- Narayanan, S., P.t Manchanda and P.K. Chintagunta (2005): Temporal Differences in the Role of Marketing Communication in New Product Categories, *Journal of Marketing Research* 42(3), 278-290.
- Newey, W.K. (1995): Convergence Rates for Series Estimators. In: Maddala, G.S., Phillips, P.C.B., Srinivasan, T.N. (Eds.), Statistical Methods of Economics and Quantitative Economics: Essays in Honor of C.R. Rao. Cambridge, MA: Blackwell
- Newey, W.K. (1997): Convergence Rates and Asymptotic Normality for Series Estimators. *Journal of Econometrics* 79, 147-168.
- Newey, W.K., and D. McFadden (1994): Large Sample Estimation and Hypothesis Testing, *Handbook of Econometrics* 4, 2111-2245.

- Osinga, E.C., P.S.H. Leeflang and J.E. Wieringa (2010): Early Marketing Matters: A Time-Varying Parameter Approach to Persistence Modeling, *Journal of Marketing Research* 47(1), 173-185.
- Özgür, O. (2011): Local Interactions, *Handbook of Social Economics*, 587-643.
- Pagan, A. and A. Ullah (1999): Nonparametric Econometrics. Cambridge UK: Cambridge University Press.
- Parker, P.M. (1992): Price Elasticity Dynamics over the Adoption Life Cycle, *Journal of Marketing Research* 29(3), 358-367.
- Parsons, L.J. (1975): The Product Life Cycle and Time-Varying Advertising Elasticities, *Journal of Marketing Research* 12(4), 476-480.
- Patacchini, E., and Y. Zenou (2012): Juvenile Delinquency and Conformism, *Journal of Law, Economics, and Organization* 28(1), 1-31.
- Perrigne, I., and Q. Vuong (2011): Nonlinear Pricing in Yellow Pages, Working Paper.
- Podolny, J.M. (2005): Status Signals: A Sociological Study of Market Competition. Princeton, NJ: Princeton University Press.
- Powell, J. L. (1986): Symmetrically Trimmed Least Squares Estimation for Tobit Models, *Econometrica*, 54, 1435-1460.
- Powell, L.M., J.A. Taurasb, and H. Rossc (2005): The Importance of Peer Effects, Cigarette Prices and Tobacco Control Policies for Youth Smoking Behavior, *Journal of Health Economics* 24(5), 950-968.
- Ray, D., and A. Robson (2012): Status, Intertemporal Choice, and Risk-Taking, *Econometrica* 80(4), 1505-1531.

- Robinson, P.M. (1988): Root-N-Consistent Semiparametric Regression, *Econometrica* 56(4), 931- 954.
- Robson, A.J. (1992): Status, the Distribution of Wealth, Private and Social Attitudes to Risk, *Econometrica* 60(4), 837-857.
- Sacerdote, B. (2001): Peer Effects with Random Assignment: Results for Dartmouth Roommates, *Quarterly Journal of Economics* 116(2), 681-704.
- Sauder, M., F. Lynn, and J.M. Podolny (2012): Status: Insights from Organizational Sociology, *Annual Review of Sociology* 38, 267-283.
- Seim, K. (2006): An Empirical Model of Firm Entry with Endogenous Product-type Choices, *RAND Journal of Economics* 37(3), 619-640.
- Shankar, V. and B.L. Bayus (2003): Network Effects and Competition: An Empirical Analysis of the Video Games Industry, *Strategic Management Journal* 24(4), 375-384.
- Shen, X., and W. Wong (1994): Convergence Rate of Sieve Estimates, *Annals of Statistics* 22(2), 580-615.
- Shen, X. (1997): On Methods of Sieves and Penalization, *Annals of Statistics* 25(6), 2555-2591.
- Sheng, S. (2012): Identification and Estimation of Network Formation Games, Working Paper.
- Steenkamp, J.E.M., V.R. Nijs, D.M. Hanssens and M.G. Dekimpe (2005): Competitive Reactions to Advertising and Promotion Attacks, *Marketing Science* 24(1), 35-54.

- Sun, Y., R.J. Carroll, and D. Li (2009): Semiparametric Estimation of Fixed-Effects Panel Data Varying Coefficient Models, *Advances in Econometrics* 25, 101-129.
- Tamer, E. (2003): Incomplete Simultaneous Discrete Response Model with Multiple Equilibria, *Review of Economic Studies* 70, 147-165.
- Trogdon, J.G, J. Nonnemaker, and J. Pais (2008): Peer Effects in Adolescent Overweight, *Journal of Health Economics* 27(5), 1388-1399.
- Van H., H.J., C.F. Mela and P. Manchanda (2004): The Dynamic Effect of Innovation on Market Structure, *Journal of Marketing Research* 41(2), 166-183.
- Wasserman, S., and K. Faust (1994): Social Network Analysis: Methods and Applications. Cambridge, UK: Cambridge University Press.
- Xu, H. (2011): Social Interactions: A Game Theoretical Approach, Working Paper.

APPENDIX

A Lemma

A.1 Lemma on Continuous S

From the definition of $\Delta_i(W_n)$, we have

$$\Delta_i(W_n) = \beta(X_i) + \sum_{j \in F_i} \alpha(0, S_i, S_j) + \sum_{j \in F_i} \gamma(S_i, S_j) \cdot \mathbb{E}[Y_j | W_n]$$

Denote information subsets $I_{\text{sub}} = \{X_i = x, S_i = s, N_{F_i} = m, \{S_j = s_j : \forall j \in F_i\}\}$ and

$I_{\text{subb}} = \{X_i = x, S_i = s, N_{F_i} = m\}$. Because $I_{\text{subb}} \subseteq I_{\text{sub}} \subseteq W_n$, we have

$$\mathbb{E}[\Delta_i(W_n) | I_{\text{sub}}] = \beta(X_i) + \sum_{j \in F_i} \alpha(0, s, s_j) + \sum_{j \in F_i} \gamma(s, s_j) \cdot \mathbb{E}[Y_j | I_{\text{sub}}]$$

$$\mathbb{E}[\Delta_i(W_n) | I_{\text{subb}}] = \beta(X_i) + \sum_{j \in F_i} \alpha(0, s, s_j) + \sum_{j \in F_i} \gamma(s, s_j) \cdot \mathbb{E}[Y_j | I_{\text{subb}}]$$

Therefore

$$\mathbb{E}[\Delta_i(W_n) | I_{\text{sub}}] - \mathbb{E}[\Delta_i(W_n) | I_{\text{subb}}] = \sum_{j \in F_i} \gamma(s, s_j) \cdot \{\mathbb{E}[Y_j | I_{\text{sub}}] - \mathbb{E}[Y_j | I_{\text{subb}}]\}$$

Without loss of generality, we denote the set of friends of individual i as $F_i = \{i_1, i_2, \dots, i_{N_{F_i}}\}$ and $s_{i_j} \neq s_{i_k} \forall i_j \neq i_k, i_j, i_k \in F_i$. We make the following rank condition.

Assumption 11. *For any individuals i , with m friends, there is variation in the conditional expectations of friends' choices, i.e.*

$$\det[Z_i^* \cdot Z_i^{*'}] \neq 0$$

where $Z_i^* = (\mathbb{E}[Y_{i_1} | I_{\text{sub}}] - \mathbb{E}[Y_{i_1} | I_{\text{subb}}], \dots, \mathbb{E}[Y_{i_1} | I_{\text{sub}}] - \mathbb{E}[Y_{i_1} | I_{\text{subb}}])$.

Assumption 11 first restricts that the status of friends provides more information for expectation formation. Second, expectations of friends' choice are linear independent when status is different. Thus we have

Lemma 4. With assumptions 1 to 3 and 11 hold, we have our model identified when S is continuous.

Proof. Through above discussion, we easily obtain the identification of $\gamma(s, s_j)$ by similar arguments as in the identification of binary status and the rank condition in assumption 11. With arbitrary s and s_j , we identify $\gamma(\cdot, \cdot)$. Denote

$$\tilde{\Delta}_i(I_{sub}) \equiv \Delta_i(I_{sub}) - \sum_{j \in F_i} \gamma(s, s_j)$$

$\tilde{\Delta}_i(I_{sub})$ is identified from the identifications of $\Delta_i(W_n)$ and γ . Therefore we have

$$\tilde{\Delta}_i(I_{sub}) \equiv \beta(x) + \sum_{j \in F_i} \alpha(0, s, s_j)$$

We here provide the identification of β and α through individuals with one or two friends. The result can be easily generalized to large but less than M friends case.

$$\mathbb{E}[\tilde{\Delta}_i(I_{sub}) | X_i = x, S_i = s, NF_i = 1, \{S_j = s' : \forall j \in F_i\}] = \beta(x) + \alpha(0, s, s')$$

$$\mathbb{E}[\tilde{\Delta}_i(I_{sub}) | X_i = x, S_i = s, NF_i = 1, \{S_j = s'' : \forall j \in F_i\}] = \beta(x) + \alpha(0, s, s'')$$

$$\begin{aligned} \mathbb{E}[\tilde{\Delta}_i(I_{sub}) | X_i = x, S_i = s, NF_i = 2, \{(S_j, S_{j'}) = (s', s'') : \forall j, j' \in F_i\}] \\ = \beta(x) + \alpha(0, s, s') + \alpha(0, s, s'') \end{aligned}$$

We achieve the identification of $\beta(x)$ by

$$\begin{aligned}
\beta(x) = & \mathbb{E}[\tilde{\Delta}_i(I_{sub})|X_i = x, S_i = s, NF_i = 1, \{S_j = s': \forall j \in F_i\}] \\
& + \mathbb{E}[\tilde{\Delta}_i(I_{sub})|X_i = x, S_i = s, NF_i = 1, \{S_j = s'': \forall j \in F_i\}] \\
& - \mathbb{E}[\tilde{\Delta}_i(I_{sub})|X_i = x, S_i = s, NF_i = 2, \{(S_j, S_{j'}) = (s', s''): \forall j, j' \in F_i\}]
\end{aligned}$$

and $\alpha(0, s, s')$ by

$$\begin{aligned}
\alpha(0, s, s') = & \mathbb{E}[\tilde{\Delta}_i(I_{sub})|X_i = x, S_i = s, NF_i = 2, \{(S_j, S_{j'}) = (s', s''): \forall j, j' \in F_i\}] \\
& - \mathbb{E}[\tilde{\Delta}_i(I_{sub})|X_i = x, S_i = s, NF_i = 1, \{S_j = s'': \forall j \in F_i\}]
\end{aligned}$$

Because s and s' are arbitrarily chosen, we have $\beta(\cdot)$ and $\alpha(0, \cdot, \cdot)$ identified.

A.2 Lemma 5

Define $N_{(i,h)}$ as the set of all friends of individual i with largest social distance h , where distance is defined as the least link number of transmission from individual i to j . Further, let $W_n^{(i,h)}$ be the information set within $N_{(i,h)}$.

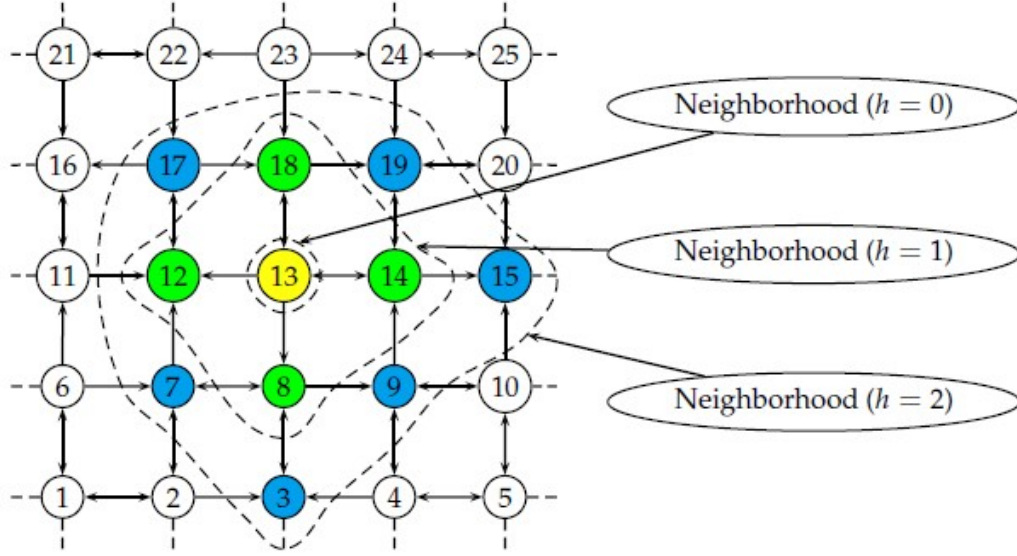


Figure 12 Social Distance and Neighborhood

Lemma 5. *With positive choice probability and smoothness, $q_i(Y_i, W_n; \theta, P_n)$ is a bounded continuous function in θ . Since Θ is compact, then $\mathcal{F}_n \equiv \{q_i(Y_i, W_n; \theta, P_n) : \theta \in \Theta\}$ can be covered by a finite number of ϵ -brackets. To apply the classical Glivenko-Cantelli argument, it suffices to show the pointwise Law of Large Number (LLN), i.e. for any $\theta \in \Theta$*

$$Q_n(\theta, P_n) - Q(\theta, P_n; n) \xrightarrow{p} 0$$

Proof. Because

$$\begin{aligned}
\mathbb{E}\{[Q_n(\theta, P_n) - Q(\theta, P_n; n)]^2\} &= \mathbb{E}\left\{\left[\frac{1}{n}\sum_{i=1}^n (q_i(Y_i, W_n; \theta, P_n) - \mathbb{E}[q_i(Y_i, W_n; \theta, P_n)])\right]^2\right\} \\
&= \frac{1}{n^2} \mathbb{E}\left\{\mathbb{E}\left(\left[\sum_{i=1}^n (q_i(Y_i, W_n; \theta, P_n) - \mathbb{E}[q_i(Y_i, W_n; \theta, P_n)|W_n])\right]^2 \middle| W_n\right)\right\} \\
&+ \frac{1}{n^2} \mathbb{E}\left\{\mathbb{E}\left(\sum_{i=1}^n (\mathbb{E}[q_i(Y_i, W_n; \theta, P_n)|W_n] - \mathbb{E}[q_i(Y_i, W_n; \theta, P_n)])^2 \middle| W_n\right)\right\} \quad (1)
\end{aligned}$$

We suppress all zero terms in RHS of equation (11). Conditional on W_n , $\{Y_i\}_{i=1}^n$ is independent with each other. Then $\{q_i(Y_i, W_n; \theta, P_n)\}_{i=1}^n$ are also conditionally independent. From the definition of q_i , we know it is bounded and continuous, therefore we have

$$\begin{aligned}
&\frac{1}{n^2} \mathbb{E}\left\{\mathbb{E}\left(\left[\sum_{i=1}^n (q_i(Y_i, W_n; \theta, P_n) - \mathbb{E}[q_i(Y_i, W_n; \theta, P_n)|W_n])\right]^2 \middle| W_n\right)\right\} \\
&= \frac{1}{n^2} \mathbb{E}\left\{\sum_{i=1}^n \mathbb{E}(q_i(Y_i, W_n; \theta, P_n) - \mathbb{E}[q_i(Y_i, W_n; \theta, P_n)|W_n]|W_n)^2\right\} \\
&= \frac{1}{n^2} \times O(n) = o(1)
\end{aligned}$$

With similar argument in the proof of Theorem 1, we have, for fixed natural number $h^* \in \mathbb{N}$,

$$\sup_{\theta \in \Theta} \left| p_i^*(W_n; \theta) - p_i^{(i, h^*)}(W_n^{(i, h^*)}; \theta) \right| \leq 2\lambda^{h^*} \quad (2)$$

where $p_i^{(i, h^*)}(W_n^{(i, h^*)}; \theta)$ is the equilibrium choice probability for individual i in the subnetwork comprised with i and her h^* -distance friends. The property in equation (2) is

called as “pairwise stability” or “network stability”, e.g., in Jackson and Wolinsky (1996); Xu (2011). Network stability is crucial for large social network analysis.

Furthermore, by Taylor expansion, we have

$$\begin{aligned} \sup_{\theta \in \Theta} & \left| p_i^*(W_n; \theta) \ln p_i^*(W_n; \theta) - p_i^{(i, h^*)}(W_n^{(i, h^*)}; \theta) \ln p_i^{(i, h^*)}(W_n^{(i, h^*)}; \theta) \right| \\ & \leq 2(1 + \ln \underline{p}) \lambda^{h^*} \end{aligned}$$

where \underline{p} is the lower bound of all choice probabilities.

Hence

$$\sup_{\theta \in \Theta} \left| \mathbb{E}[q_i(Y_i, W_n; \theta, P_n) | W_n] - \mathbb{E}[q_i(Y_i, W_n; \theta, P_n) | W_n^{(i, h^*)}] \right| \leq 8(1 + \ln \underline{p}) \lambda^{h^*} \quad (3)$$

Because

$$\begin{aligned} & \mathbb{E} \left\{ \left(\sum_{i=1}^n [\mathbb{E}(q_i(Y_i, W_n; \theta, P_n) | W_n) - \mathbb{E}(q_i(Y_i, W_n; \theta, P_n))] \right)^2 \right\} \\ &= \mathbb{E} \left\{ \left(\sum_{i=1}^n [\mathbb{E}(q_i(Y_i, W_n; \theta, P_n) | W_n) - \mathbb{E}(q_i(Y_i, W_n; \theta, P_n) | W_n^{(i, h^*)})] \right)^2 \right\} \\ &+ \mathbb{E} \left\{ \left(\sum_{i=1}^n [\mathbb{E}(q_i(Y_i, W_n; \theta, P_n) | W_n^{(i, h^*)}) - \mathbb{E}(q_i(Y_i, W_n; \theta, P_n))] \right)^2 \right\} \end{aligned}$$

With equation (13), we have

$$\mathbb{E} \left\{ \left(\sum_{i=1}^n [\mathbb{E}(q_i(Y_i, W_n; \theta, P_n) | W_n) - \mathbb{E}(q_i(Y_i, W_n; \theta, P_n) | W_n^{(i, h^*)})] \right)^2 \right\} = O(n^2 \lambda^{2h^*})$$

Furthermore

$$\begin{aligned}
& \mathbb{E} \left\{ \left(\sum_{i=1}^n [\mathbb{E}(q_i(Y_i, W_n; \theta, P_n) | W_n^{(i, h^*)}) - \mathbb{E}(q_i(Y_i, W_n; \theta, P_n))] \right)^2 \right\} \\
&= \sum_{i=1}^n \sum_{j \in N_{(i, h^*)}} \text{Cov} \left(\mathbb{E}(q_i(Y_i, W_n; \theta, P_n) | W_n^{(i, h^*)}), \mathbb{E}(q_j(Y_j, W_n; \theta, P_n) | W_n^{(j, h^*)}) \right) \\
&+ \sum_{i=1}^n \text{Var} \left(\mathbb{E}(q_i(Y_i, W_n; \theta, P_n) | W_n^{(i, h^*)}) \right) = O(nM^{h^*}) + O(n)
\end{aligned}$$

Choose $h^* = \frac{b \ln n}{\ln M}$ for some $b \in (0, 1)$. Then $h^* \rightarrow \infty$ as $n \rightarrow \infty$ and $M^{h^*} = o(n)$.

h^* only serves in this proof. Hence

$$\mathbb{E} \left\{ \left(\sum_{i=1}^n [\mathbb{E}(q_i(Y_i, W_n; \theta, P_n) | W_n^{(i, h^*)}) - \mathbb{E}(q_i(Y_i, W_n; \theta, P_n))] \right)^2 \right\} = O(n)$$

Therefore we conclude

$$\begin{aligned}
& \mathbb{E} \left\{ \left(\sum_{i=1}^n [\mathbb{E}(q_i(Y_i, W_n; \theta, P_n) | W_n) - \mathbb{E}(q_i(Y_i, W_n; \theta, P_n))] \right)^2 \right\} = o(1) + \frac{O(n^2 \lambda^{2h^*} + n)}{n^2} \\
&= o(1)
\end{aligned} \tag{4}$$

where $\lambda^{2h^*} \rightarrow 0$. Therefore with equations (2) to (4), we have

$$\mathbb{E}[Q_n(\theta, P_n) - Q(\theta, P_n; n)]^2 \rightarrow 0 \Rightarrow Q_n(\theta, P_n) \xrightarrow{p} Q(\theta, P_n; n)$$

The pointwise LLN obtained.

B Proofs

B.1 Proof of Theorem 1

Proof. We show the uniqueness by contradiction. For notation simplification, denote $p_i = p_i(W_n; \theta)$ and $P = (p_1, \dots, p_n)'$ as the vector of equilibrium choice probabilities. Suppose there are two equilibria $P^{(1)}$ and $P^{(2)}$, $P^{(1)} \neq P^{(2)}$. Let $\gamma(S_i, S_j) = \alpha(1, S_i, S_j) - \alpha(0, S_i, S_j)$. For any $i, j \in I$, let

$$\Gamma_i(W_n, \{p_j\}_{j \in F_i}; \theta) = \frac{\exp\{\beta(X_i) + \sum_{j \in F_i} [\alpha(0, S_i, S_j) + \gamma(S_i, S_j) \cdot p_j]\}}{1 + \exp\{\beta(X_i) + \sum_{j \in F_i} [\alpha(0, S_i, S_j) + \gamma(S_i, S_j) \cdot p_j]\}}$$

Therefore

$$p_i = \Gamma_i(W_n, \{p_j\}_{j \in F_i}; \theta)$$

We assume there is at least one loop connection. With loss of generality, take i as one individual in this loop.

$$\begin{aligned} |p_i^{(1)} - p_i^{(2)}| &= \left| \sum_{j \in F_i} \Gamma_i(W_n, \{p_j^+\}_{j \in F_i}; \theta) \cdot [1 - \Gamma_i(W_n, \{p_j^+\}_{j \in F_i}; \theta)] \gamma(S_i, S_j) \cdot (p_j^{(1)} \right. \\ &\quad \left. - p_j^{(2)}) \right| \end{aligned}$$

where p_j^+ is a choice probability between $p_j^{(1)}$ and $p_j^{(2)}$. Because $\Gamma_i \in (0, 1)$, we have $\Gamma_i(W_n, \{p_j^+\}_{j \in F_i}; \theta) \cdot [1 - \Gamma_i(W_n, \{p_j^+\}_{j \in F_i}; \theta)] \leq \frac{1}{4}$. With assumption 2 and the definition of λ , we have

$$|p_i^{(1)} - p_i^{(2)}| < \frac{M}{4} \max_{S_i, S_j \in \{L, H\}} |\gamma(S_i, S_j)| \times \max_{j \in F_i} |p_j^{(1)} - p_j^{(2)}| \leq \lambda \max_{j \in I} |p_j^{(1)} - p_j^{(2)}|$$

Therefore we have

$$\max_{i \in I} |p_i^{(1)} - p_i^{(2)}| < \lambda \max_{j \in I} |p_j^{(1)} - p_j^{(2)}|$$

which contradicts with $\lambda < 1$ in assumption 3. This completes our proof of the uniqueness.

B.2 Proof of Theorem 3

Proof. Clearly, it suffices to show that

$$\begin{aligned} & \mathbb{E}\{Y_i \ln \Gamma_i(P_n(\theta_0), W_n; \theta_0) + (1 - Y_i)[1 - \ln \Gamma_i(P_n(\theta_0), W_n; \theta_0)]\} \\ & \geq \mathbb{E}\{Y_i \ln \Gamma_i(P_n(\theta_0), W_n; \theta) + (1 - Y_i)[1 - \ln \Gamma_i(P_n(\theta_0), W_n; \theta)]\} \end{aligned}$$

Note that by definition $P_n(\theta)$ solves $P_n = \Gamma(\theta, P_n)$ and the latter admits a unique solution P_n^* by Theorem 1. We have that $P_n(\theta_0) = P_n^*$. Hence, it is equivalent to show that

$$\theta_0 = \arg \max_{\theta \in \Theta} \mathbb{E}\{Y_i \ln \Gamma_i(P_n^*, W_n; \theta) + (1 - Y_i)[1 - \ln \Gamma_i(P_n^*, W_n; \theta)]\}$$

which holds by the fact that $Y_i \ln \Gamma_i(P_n^*, W_n; \theta) + (1 - Y_i)[1 - \ln \Gamma_i(P_n^*, W_n; \theta)]$ is the underlying log-likelihood and that the regular conditions in MLE hold.

B.3 Proof of Theorem

Proof. The proof is similar as that in Aguirregabiria and Mira (2007); Newey and McFadden (1994). With assumption 8, we have that $\theta_{\text{NPL}} = \theta_0$. Recall that the pseudo likelihood function is $Q_n(\theta, P_n)$ in NPL estimation. Denote Λ_0 and Λ_n as the fixed points set for $Q(\theta, P_n; n)$ and $Q_n(\theta, P_n)$ respectively. Define the function

$$T(\theta, P_n) \equiv \max_{c \in \Theta} \{Q(c, P_n; n)\} - Q(\theta, P_n; n)$$

Because $Q(\theta, P_n; n)$ is continuous and $\Theta \times [0,1]^n$ is compact, Berge's maximum theorem establishes that $T(\theta, P_n)$ is a continuous function. By construction, $T(\theta, P_n) \geq 0$ for any (θ, P_n) . Let \mathcal{E} be the set of vectors (θ, P_n) that are fixed points of the equilibrium mapping Γ , i.e.,

$$\mathcal{E} \equiv \{(\theta, P_n) \in \Theta \times [0,1]^n : P_n = \Gamma(\theta, P_n)\}$$

Given that $\Theta \times [0,1]^n$ is compact and Γ is continuous, then \mathcal{E} is a compact set. By definition, the set Λ_0 is included in \mathcal{E} . Let $B_\epsilon(\theta_0) = \{\theta \in \mathbb{R}^L : \|\theta - \theta_0\| < \epsilon, \forall \epsilon > 0\}$ be an arbitrarily small open ball that contains θ_0 , we have that $B_\epsilon(\theta_0) \cap \mathcal{E}$ is also compact. Define the constant

$$\tau = \min_{(\theta, p_1) \in B_\epsilon^c(\theta_0) \cap \mathcal{E}} T(\theta, P_n) > 0$$

Define the event

$$A_n \equiv \left\{ |Q_n(\theta, P_n) - Q(\theta, P_n; n)| < \frac{\tau}{2} \text{ for all } (\theta, P_n) \in \Theta \times [0,1]^n \right\}$$

Let $(\theta^{(n)}, P_n^{(n)})$ be an element of Λ_n . Then we have that A_n implies

$$G(\theta^{(n)}, P_n^{(n)}; n) > G_n(\theta^{(n)}, P_n^{(n)}) - \frac{\tau}{2}$$

and

$$G_n(\theta^{(n)}, P_n^{(n)}) > G(\theta^{(n)}, P_n^{(n)}; n) - \frac{\tau}{2}$$

Furthermore, we have $G_n(\theta^{(n)}, P_n^{(n)}) \geq G_n(\theta, P_n^{(n)})$ from the NPL fixed point definition.

Therefore we have that $G(\theta^{(n)}, P_n^{(n)}; n) > G(\theta, P_n^{(n)}; n) - \tau$. Thus we have the following derivation:

$$\begin{aligned}
A_n &\Rightarrow \left\{ G(\theta^{(n)}, P_n^{(n)}; n) > G(\theta, P_n^{(n)}; n) - \tau \text{ for any } \theta \in \Theta \right\} \\
&\Rightarrow \left\{ G(\theta^{(n)}, P_n^{(n)}; n) > \max_{\theta \in \Theta} G(\theta, P_n^{(n)}; n) - \tau \right\} \\
&\Rightarrow \left\{ \tau > T(\theta^{(n)}, P_n^{(n)}; n) \right\} \Rightarrow \min_{(\theta, p_1) \in B_\epsilon^c(\theta_0) \cap \mathcal{E}} T(\theta, P_n) > T(\theta^{(n)}, P_n^{(n)}; n) \\
&\Rightarrow \left\{ (\theta^{(n)}, P_n^{(n)}) \in B_\epsilon(\theta_0) \right\}
\end{aligned}$$

The last induction uses the fact that $(\theta^{(n)}, P_n^{(n)}) \in \mathcal{E}$. Therefore $\Pr(A_n) \leq \Pr\{(\theta_n, p_{1(n)}) \in B_{\epsilon}(\theta_0)\}$. Because $\Pr(A_n) \rightarrow 1$ as $n \rightarrow \infty$ from appendix A.1, and ϵ is an arbitrarily small constant, we have

$$(\theta^{(n)}, P_n^{(n)}) \xrightarrow{p} (\theta_0, P_n^*)$$

From the definition of Λ_n , we have that $\hat{\theta}_{NPL} \xrightarrow{p} \theta_0$.

B.4 Proof of Theorem 4

Proof. From the first order condition we have that

$$\frac{\partial Q_n(\theta, P_n)}{\partial \theta} \Big|_{(\theta, P_n) = (\hat{\theta}_{NPL}, \hat{P}_n)} = 0$$

For notational simplicity, we denote $\frac{\partial Q_n(\theta, P_n)}{\partial \theta} \Big|_{(\theta, P_n) = (\hat{\theta}_{NPL}, \hat{P}_n)} = \frac{\partial Q_n(\hat{\theta}, \hat{P}_n)}{\partial \theta}$. Taking Taylor expansion of above equation around the true parameter (θ_0, P_n^*) , we have

$$\begin{aligned}
&\frac{\partial Q_n(\theta_0, P_n^*)}{\partial \theta} + \frac{\partial^2 Q_n(\theta_0, P_n^*)}{\partial \theta \partial \theta'} (\hat{\theta}_{NPL} - \theta_0) + \frac{\partial^2 Q_n(\theta_0, P_n^*)}{\partial \theta \partial P_n} \frac{\partial P_n}{\partial \theta} (\hat{\theta}_{NPL} - \theta_0) + o_p(1) \\
&= 0
\end{aligned} \tag{5}$$

where the $o_p(1)$ comes from similar argument in the proof of lemma 5. Further, from the definition of Q_n , we have

$$\frac{\partial Q_n(\theta_0, P_n^*)}{\partial \theta} = \frac{1}{n} \sum_{i=1}^n \frac{\partial q_i(Y_i, W_n; \theta_0, P_n^*)}{\partial \theta} = \frac{1}{n} \sum_{i=1}^n X_i^*(Y_i - p_i^*(W_n; \theta_0)) \quad (6)$$

From equation (5) and equation (6), we have that

$$\begin{aligned} & \left[\frac{\partial^2 Q_n(\theta_0, P_n^*)}{\partial \theta \partial \theta'} + \frac{\partial^2 Q_n(\theta_0, P_n^*)}{\partial \theta \partial P_n} \frac{\partial P_n}{\partial \theta} \right] \sqrt{n}(\hat{\theta}_{NPL} - \theta_0) \\ &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i^*(Y_i - p_i^*(W_n; \theta_0)) + o_p(\sqrt{n}) \end{aligned} \quad (7)$$

For any $\kappa \in \mathbb{R}^{d+d_s}$, let $\rho(\kappa, W_n) = \kappa' \left[\frac{\partial Q_n(\theta_0, P_n^*)}{\partial \theta} \right]^{-\frac{1}{2}}$, then there is

$$\frac{1}{\sqrt{n\kappa'\kappa}} \rho(\kappa, W_n) \sum_{i=1}^n X_i^*(Y_i - p_i^*(W_n; \theta_0)) \xrightarrow{d} N(0, 1) \quad (8)$$

To motivate the condition of Cramer-Wold Theorem, we rewrite equation (8) as

$$\kappa' \left[\left[\frac{\partial Q_n(\theta_0, P_n^*)}{\partial \theta} \right]^{-\frac{1}{2}} \sum_{i=1}^n X_i^*(Y_i - p_i^*(W_n; \theta_0)) \right] \xrightarrow{d} \frac{\kappa'}{\sqrt{\kappa'\kappa}} N(0, I_{d+d_s})$$

Because κ is an arbitrary vector, by Cramer-Wold Theorem, we have

$$\left[\left[\frac{\partial Q_n(\theta_0, P_n^*)}{\partial \theta} \right]^{-\frac{1}{2}} \sum_{i=1}^n X_i^*(Y_i - p_i^*(W_n; \theta_0)) \right] \xrightarrow{d} N(0, I_{d+d_s})$$

Therefore we have

$$\sqrt{n}(\hat{\theta}_{NPL} - \theta_0) \xrightarrow{d} N(0, \Omega(\theta_0))$$

where

$$\Omega(\theta_0) = V^{-1}(\theta_0) \cdot V_1(\theta_0) \cdot V^{-1}(\theta_0).$$

C Smooth Varying Coefficient Model

$$y_{it} = \alpha(Z_{1it}, Z_{2it}) + X'_{it}\beta(Z_{1it}, Z_{2it}) + \varepsilon_{it}$$

Where Z_{1it} is the continuous smooth variable and Z_{2it} is the discrete smooth variable. X_{it} is the regressor. In our market response context, X_{it} contains marketing mix variables, advertising and detailing, and competitive marketing expenditure. Z_{1it} is the time in the market and Z_{2it} is the order of entry. Denote $Z_{it} = (Z_{1it}, Z_{2it})$, $W_{it} = (1, X'_{it})'$ and $\theta(Z_{it}) = (\alpha(Z_{it}), \beta(Z_{it})')'$. Therefore we have

$$y_{it} = W'_{it}\theta(Z_{it}) + \varepsilon_{it}$$

$$\Rightarrow E(W_{it}y_{it}|Z_{it}) = E(W_{it}W'_{it}|Z_{it})\theta(Z_{it}) + 0$$

$$\Rightarrow \theta(Z_{it}) = [E(W_{it}W'_{it}|Z_{it})]^{-1}E(W_{it}y_{it}|Z_{it})$$

Replace the conditional expectations with their nonparametric kernel estimators we have

$$\hat{\theta}(Z_{it}) = [\hat{E}(W_{it}W'_{it}|Z_{it})]^{-1} \hat{E}(W_{it}y_{it}|Z_{it}),$$

where

$$\hat{E}(W_{it}W'_{it}|Z_{it} = z) = \frac{\sum_{j=1}^n \sum_{s=1}^T W_{js}W'_{js}K\left(\frac{Z_{js} - z}{h}\right)}{\sum_{j=1}^n \sum_{s=1}^T K\left(\frac{Z_{js} - z}{h}\right)}$$

$$\hat{E}(W_{it}y_{it}|Z_{it} = z) = \frac{\sum_{j=1}^n \sum_{s=1}^T W_{js}y_{js}K\left(\frac{Z_{js} - z}{h}\right)}{\sum_{j=1}^n \sum_{s=1}^T K\left(\frac{Z_{js} - z}{h}\right)}$$

where

$$K\left(\frac{Z_{js} - z}{h}\right) = K_1\left(\frac{Z_{1js} - z_1}{h_1}\right) \times L_2\left(\frac{Z_{2js} - z_2}{h_2}\right)$$

Kernel function takes standard normal form when Z is continuous

$$K_1\left(\frac{Z_{1js} - z_1}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(Z_{1js} - z_1)^2}{h^2}}$$

Kernel function takes the following form when Z is discrete

$$L_2\left(\frac{Z_{2js} - z_2}{\lambda}\right) = \begin{cases} 1 & \text{if } Z_{2js} = z_2 \\ \lambda & \text{if } Z_{2js} \neq z_2 \end{cases}$$

where $\lambda \in [0,1]$ is the bandwidth for the discrete choice variable Z_2 and h is the bandwidth for the continuous smooth variable respectively.

The semiparametric estimator for varying coefficients are consistent, $\hat{\theta}(Z_{it}) - \theta_0(Z_{it}) = o_p(1)$. This result can be found in Cai and Li (2008). In our smooth coefficient model, we do not incorporate the fixed effects since the order of entries identifies the brands and are cross sectional though we put a time subscript in them. The results for smooth coefficient panel data model with fixed effects can be found in Sun, Carroll and Li (2009).

In our application, the Cross-Validation Least Square method for bandwidth selection under-smooths the coefficient estimation. Therefore we adopt ad hoc bandwidth selection in our analysis. The ad hoc bandwidth selection is widely used in the empirical application of semiparametric and nonparametric methods, see Li and Racine (2007) and Horowitz (2009) for details. The ad hoc bandwidth selections are as follows

$$h_{ad} = c_1 \text{std}(Z_{1it}) \times nt^{-\frac{1}{q+4}}$$

$$\lambda_{\text{ad}} = c_2 \text{std}(Z_{2it}) \times n t^{-\frac{2}{q+4}}$$

Where q is the dimension of Z_{it} . c_1 and c_2 are positive constant, which we take as ones.